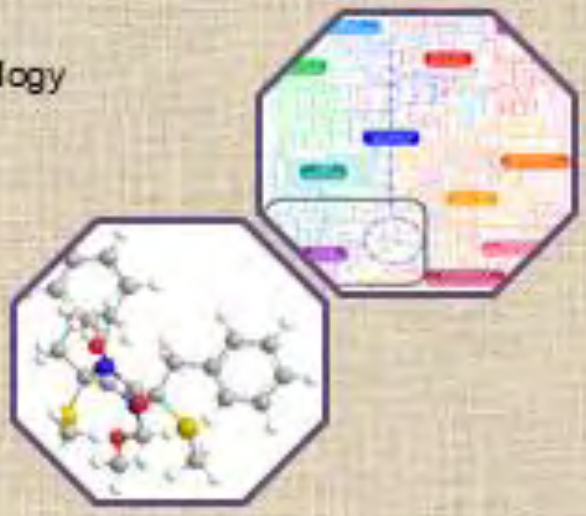
University of Thessaly
Department of Biochemistry and Biotechnology
Master in Toxicology



Exposomics analysis linking environmental exposures to neurodevelopmental disorders:

A combination of metabolomics and bioinformatics analysis.

Tsioka Aikaterini

Thessaloniki 2017

"Exposomics analysis linking environmental exposures to neurodevelopmental disorders: A combination of metabolomics and bioinformatics analysis"

Aikaterini Tsioka

Biologist, Postgraduate Student of Toxicology

Three-member Advisory Committee:

Dimosthenis A. Sarigiannis: M.S., PhD is Associate Professor specializing on environment and health issues at the Department of Chemical Engineering of the Aristotle University of Thessaloniki, Visiting Professor at the Master's Program on Toxicology and Environmental Risk at the Medical School of the University of Pavia and senior scientist at the Chemical Assessment and Testing unit of the Institute for Health and Consumer Protection at the European Commission's Joint Research Centre.

Spyridon Karakitsios: Head of Exposure Biology Unit, scientific collaborator in EnvE-Lab in Department of Chemical Engineering of the Aristotle University of Thessaloniki. Dr. Karakitsios is an environmental healthscientist, with studies in physics (degree), environmental and computational chemistry (M.Sc.) and applied biology (PhD) of the University of Ioannina. His occupational field is biochemical and environmental engineering.

Andreas K. Tsakalof: Associate Professor of Medical Chemistry, Laboratory of Biochemistry, Faculty of Medicine University of Thessaly. His research activities on: fabrication of molecularly imprinted polymers and their application for the identification and isolation of bioactive molecules for the regulation of Hypoxia Inducible Factors (HIFs) and bio-molecules involved in HIFs activation pathways and application of chromatography and chromatography-mass spectrometry in clinical chemistry and toxicology.

Contents

Preface- Acknowledgments

This thesis is the final work of my Master in Toxicology, Department of Biochemistry and Biotechnology, University of Thessaly. The study was conducted at the Environmental Engineering Laboratory, Department of Chemical Engineering in Aristotle University of Thessaloniki during the academic year 2016-2017.

I would like to thank Dr. Dimosthenis Sarigiannis, Associate Professor in Department of Chemical Engineering and Director in Environmental Engineering Laboratory (EnvELab), AUTH and Dr. Spyridon Karakitsios, Head of Exposure Biology Unit in EnvELab, Department of Chemical Engineering, AUTH for the confidence they showed me procuring this study, the time taken for ongoing help and advice and the corrections they made in my text.

Moreover, I would like to thank Dr. Andreas Tsakalof, Associate Professor in Medicine School, University of Thessaly, for his indications in my text.

I also would like to thank Dr. Aikaterini Gabriel and Dr. Konstantinos Tsarchopoulos for their assistance in the experimental part of the work with LC-MS/MS instrument and the indications in my text.

Then many thanks to Evangelos Handakas, a PhD student in EnvELab for his help with the statistical analysis and the results processing. A big thank you to Nafsika Papaioannou, an undergraduate student in EnvELab for her help with the introduction to the bioinformatics analysis. This brings me to thanks the PhD students Krystalia Papadaki and Dimitris Chapizanis, the remaining members of the EnvELab and my fellow master students for their help and support.

Sincere gratitude goes to my family, my parents Konstantinos and Antigone and my brother George for all the love, support and positive energy they provided me with during this period.

Thessaloniki 13.01.2017

Katerina Tsioka

Abstract

Introduction: Metabolomics analysis refers to the range of metabolites present in a person's body under healthy or pathological status. Metabolites can be used as biomarkers to provide mechanistic explanation of clinical observations through metabolic pathway analysis. Thus, beyond environmental, dietary and sociodemographic factors, neurodevelopmental disorders could be also associated to the perturbation of normal levels of metabolites in body fluids and the respective pathway(s).

Aim: The current study aimed at detecting metabolites from urine samples and to correlate them with neurodevelopmental disorders in children.

Material and Methods: 92 urine samples from children (6-12 years old) were collected around an industrial complex in Taranto, Italy. The metabolites were identified using LC-MS/MS. Using bioinformatics methods, the metabolic pathways which involved the detected metabolites were identified. For statistical analysis, we used the Environment-Wide Association paradigm to correlate environmental, nutritional, sociodemographic factors and induced metabolic pathways from exposure to the above with clinically observed neurodevelopmental disorders.

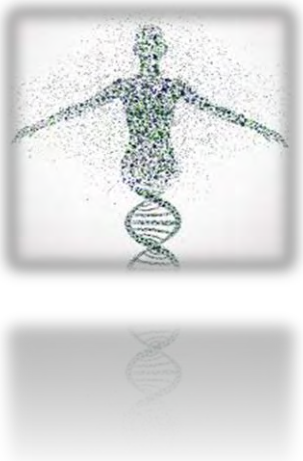
Results: Maternal and paternal educational level, as well as the overall socioeconomic status of the family were positively associated with healthy children's neurodevelopment. Positive is also the influence of Se in soil, while exposure to manganese and proximity of residence to the industrial facility have a negative impact on child neurodevelopment. Various dietary items and metabolic pathways have been associated to neurodevelopmental progress, either positively or negatively.

Conclusion: Neurodevelopmental disorders are related to environmental, nutritional, demographic factors and the metabolic pathways induced from exposure of the study subjects to the factors delineated above. Applying an environment-wide association approach would lead to the identification of the most critical parameters associated with neurodevelopmental disorders in

early life and thus provide guidance regarding the most cost-effective targeted interventions to alleviate such disorders as early as possible.

Part I

Introduction



Omics techniques

The English language neologism omics informally refers to a field of study in biology ending in –*omics*, such as genomics, proteomics or metabolomics to denote the use of high throughput analysis methods at the respective level of biological organization (genes, proteins, metabolites). The related suffix –ome is used to address the objects of study of such fields, such as the genome, proteome or metabolome respectively. Omics aims at the collective characterization and quantification of pools of biological molecules that translate into the structure, function, and dynamics of an organism or organisms. They are aimed primarily at the universal detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a specific biological sample in a non-targeted and non-biased manner(Horgan and Kenny, 2011).Genomics is the study of genomes of organisms, transcriptomics is the study of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA, produced in one or a population of cells, their structures and functions, proteomics is the study of proteins, their structures and functions and metabolomics is the systematic study of the unique chemical fingerprints that specific cellular processes leave behind. All this can also be referred to as high-dimensional biology; the integration of these techniques is called systems biology (Westerhoff and Palsson, 2004). The basic aspect of these approaches is that a complex system can be understood more thoroughly if considered as a whole. Systems biology experiments are hypothesis-generating, using holistic approaches where no hypothesis is known or prescribed but all data are acquired and analyzed to define a hypothesis that can be further tested (Kell and Oliver, 2004). A large number of applications and the usefulness in this field have been discovered. Omics technology can be applied not only for the greater understanding of normal physiological processes but also in disease processes where they play a role in screening, diagnosis and prognosis as well as aiding our understanding of the etiology of diseases(Horgan and Kenny, 2011). Through omics strategies discovery of biomarkers is possible as these techniques investigate multiple molecules simultaneously. Omics investigation is increasingly being used in drug discovery and assessment of

their toxicity and efficacy (Gerhold et al., 2002; Kell, 2006).In the future, systems biology may enable us to develop new approaches that will be predictive, preventive and personalized (Horgan and Kenny, 2011).

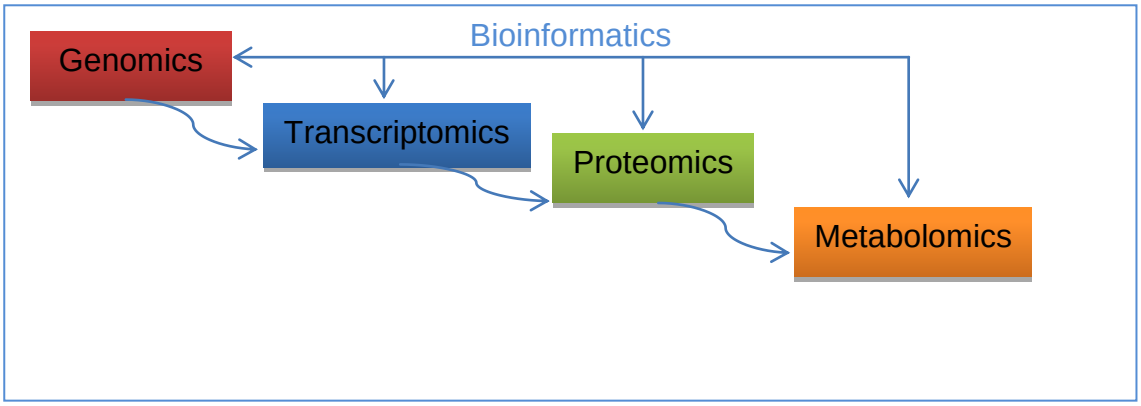


Figure 1: The hierarchy and information flow of –omics technologies.

Biomarkers

The use of the term "biomarker" dates back to as early as 1980. In 1998, the National Institutes of Health Biomarkers Definitions Working Group defined a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Strimbu and Tavel, 2010). A biomarker, or biological marker, generally refers to a measurable indicator that reflects a particular physiological state of organism. The term is also occasionally used to refer to a substance the presence of which indicates the existence of a living organism (Strachan, 2017). This material is used in many scientific fields. Biomarker measurements can help to explain empirical results of clinical trials by relating the effects of interventions on molecular and cellular pathways to clinical responses. In doing so, biomarkers provide an avenue for researchers to gain a mechanistic understanding of the differences in clinical response that may be influenced by uncontrolled variables (for example, drug metabolism).

In medicine, biomarkers are often compounds isolated from serum, urine, or other biological samples that can be used as an indicator of the presence or

severity of a particular disease state. Biomarkers can also be used to assess the effectiveness of particular therapies in ameliorating the effects of a disease. By using easily obtained and assayed biomarkers to monitor a patient's reaction to a particular drug, it is possible to determine whether treatment is effective for that individual by measuring drug response rate or toxic effects associated with the drug. This information could eventually lead to earlier detection of adverse drug response, reducing the number of costly laboratory tests – and possible other medical interventions – necessary to adjust proper dosage of a drug. Biomarkers are also important in the development of new drug therapies. Researchers can use changes in progression markers to understand if and how a new therapy is successfully slowing – or even reversing – the disease process (Feigin, 2004). Results of studies like these will allow researchers to focus efforts and resources on the most effective therapies, thus reducing the time and cost to bring a new therapy to market and eventually to the patient (Bystritsky et al., 2013).

The use of biomarkers will help us evaluate potential exposures to pesticides as well as predict effects that may result, allowing us to make decisions that are more protective of human health. Biomarkers are commonly grouped into biomarkers of exposure, effect and susceptibility. Biomarkers of exposure are used to assess the amount of a chemical that is present within the body. Many chemicals can be measured in urine, blood, saliva, and, if they are fat soluble, in body fat and breast milk (e.g., DDT). Biomarkers of effect are indicators of a change in biologic function in response to a chemical exposure. Thus, they more directly relate to insight into the potential for adverse health effects compared with biomarkers of exposure. Biomarkers of susceptibility are factors that may make certain individuals more sensitive to chemical exposure. Biomarkers of susceptibility include genetic factors that may influence how the body interacts with a chemical and other biological factors related to nutritional status, health status, lifestyle, and life stage that may affect an individual’s susceptibility to chemical exposure

Biomarkers can take many different forms, including particular proteins or peptides (e.g., prostate-specific antigen as an indicator of increased risk for prostate cancer), antibodies (e.g., anti-citrullinated protein antibodies for rheumatoid arthritis), cell types (e.g., white blood cell counts in infection or cancer), metabolites (e.g., phenylalanine in urine of newborns with phenylketonuria), lipids (e.g., cholesterol and other lipid levels in cardiovascular disease), hormones (e.g., thyroid stimulating hormone in Hashimoto's Disease), enzyme levels (e.g., various hepatic enzymes for liver cancer), physiological states such as blood pressure or fever, or imaging studies of particular organs or organ systems (e.g., neural degeneration in Parkinson's Disease). A biomarker can also be a substance introduced into a patient to assess how internal organ systems are functioning, such as radioactive iodine used to measure thyroid function. Ultimately, biomarkers can be used to detect a change in the physiological state of a patient that correlates with the risk or progression of a disease or with the susceptibility of a disease to a given treatment (Phadikar et al., 2017). Biomarkers hold great promise for personalized medicine as information gained from diagnostic or progression markers can be used to tailor treatment to the individual for highly efficient intervention in the disease process (Mandel et al., 2010).

An ideal biomarker may fill one of many different roles. A biomarker may be suitable for the early diagnosis of a disease, either as part of a routine screening exam or at the first sign of a questionable symptom (Caliendo et al., 2013). A biomarker may also appear or disappear over the course of disease progression and thus be useful in determining the prognosis of a disease within an individual. Another biomarker may change as a drug therapy is started, adjusted or discontinued, ultimately aiding in the monitoring of the patient's response to that particular therapy.

Large-scale studies are currently underway on the discovery and validation of new and more effective biomarkers. The driving force behind these studies is the revolution of "omics" medicine. Instead of looking at one potential biomarker at a time, new techniques in genomics, transcriptomics,

proteomics, metabolomics, lipidomics and glycomics have allowed investigators to identify patterns in the changes of tens, hundreds and even thousands of genes and compounds that correlate with disease state.

Biomarkers isolated from patient specimens may be proteins, polypeptides, lipids, hormones, metabolic intermediates, or nucleic acids and their derivatives. Each of these compounds require different collection, processing and storage procedures and conditions (Tuck et al., 2009). Furthermore, the time between collection of the specimen and processing for long-term storage must be minimized to limit time and temperature-dependent loss of potential biomarkers (Yin et al., 2015). For example, temperature changes of only a few degrees can induce degradation of biomarker proteins, so most investigators and large-scale studies act conservatively by choosing ultra-low freezers (-80^{o}C) for storing biofluids like serum, plasma, and urine.

Metabolomics

The beginning of metabolomics traces back all the way to 2000-1500 B.C. when traditional Chinese doctors began using ants in order to evaluate the urine of patients to determine if the urine contained the high glucose of diabetics(Wikipedia). By 1905 J.J. Thomson of the University of Cambridge developed the first mass spectrometer. Also in this year there was more work in determining which other substances were in urine and Otto Knut Olof Folin reported on methods for analysis of urine for urea, ammonia, creatinine, uric acid. His findings were all published in one issue of Physical Review. The next step in the path to modern Metabolomics came by 1946 when Felix Bloch of Stanford and Edward Purcell of Harvard simultaneously published the first NMR in the same issue of Physical Review. Arthur B Robinson and Linus Pauling, by studying early chromatographic separations in urine, they found that the chemical constituents of the urine were loaded with useful information. The first paper on Metabolomics, though not called metabolomics at the time, was by Robinson and Pauling in 1971. It was titled "*Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography*" and was published *in Proceedings of the National Academy of Sciences*. The name metabolomics was coined in the late 1990s, the first paper using the

word metabolome is Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. (1998). Systematic functional analysis of the yeast genome. In 2007 published the first draft of Human Metabolome project which consists of 2500 metabolites, 1200 drugs and 3500 food components. The group is using advanced methods in NMR spectroscopy, mass spectrometry, multi-dimensional chromatography and machine learning to facilitate this work.

By analogy with the other “omics” techniques, metabolome analysis is referred to the range of metabolites present in a person’s body at normal and pathogenic conditions. Metabolomics, which is the endpoint of the Omics cascade (Figure 1)and therefore the last step in the cascade before the phenotype, is the comprehensive study of low molecular weight metabolites, and the metabolome represents the metabolite profiles of the cellular processes in a cell, tissue, organ, or organism(Kobayashi et al., 2013).Metabolomics, or metabolome analysis, involves two approaches, metabolite profiling and metabolic fingerprinting (Dettmer et al., 2007). In metabolite profiling, the selected metabolites in a particular environment are identified, and then, a quantitative or semi-quantitative assessment is performed, although metabolic fingerprinting is used to initially examine how metabolite patterns change in response to various stimuli, for example, diseases, toxic exposure, or environmental change (Kobayashi et al., 2013).This knowledge may be a useful tool for the identification of novel biomarkers capable of being used as an early diagnostic tool. The development of metabolomics technologies is a state-of-art method with rapidly progress, which differ in their sensitivity and selectivity, but allow us to obtain a reproducible data, at relatively low cost (Lei et al., 2011).These technologies, which are typically based on nuclear magnetic resonance analysis (NMR), gas mass spectrometry (GC/MS), liquid chromatography mass spectrometry (LC/MS), and capillary electrophoresis mass spectrometry (CE/MS), have been well-documented in the literature and have been applied to various research fields including the medical field(Kobayashi et al., 2013).However, the complete set of metabolites from each sample must be extracted by a combination of these available tools. Jacod and Yemelin in

their study report that most of the publications referring to metabolome profiling in fungal pathogens are focused on the examination and screening for secondary metabolites, rather than primary metabolites. For this purpose LC-MS is the tool of choice, as in the most cases using the other methods for screening of secondary metabolites are unsuited (Tan et al., 2009).

Neurodevelopmental disorders

The spectrum of neurodevelopmental disorders includes several clinical observations such as autism, Fetal alcohol spectrum disorders, as well as poor performance in several tests (e.g. Bayley test, Wechsler Intelligence Scale for children). In recent years, the use of omics has resulted in the identification of early biological events (mainly transcriptomics and metabolomics) related to these outcomes.

With regard to metabolomics, different metabolite profiles have been identified in children with fetal alcohol spectrum disorders (FASD). The mechanisms underlying FASD are incompletely understood, and biomarkers to identify those at risk are lacking. From a metabolomics analysis of embryoid bodies and neural lineages derived from human embryonic stem cells (hES) to identify the neural secretome produced in response to ethanol (EtOH) exposure (Palmer et al., 2012). It was found that EtOH treatment induced statistically significant changes to metabolite abundance in human embryoid bodies (180 features), neural progenitors (76 features), and neurons (42 features). There were no shared significant features between different cell types. Fifteen features showed a dose-response to EtOH. Four chemical identities were confirmed: L-thyroxine, 5'-methylthioadenosine, and the tryptophan metabolites, L-kynurenine and indoleacetaldehyde. One feature with a putative annotation of succinyladenosine was significantly increased in both EtOH treatments. As a result, it was found that EtOH exposure induces statistically significant changes to the metabolome profile of human embryoid bodies, neural progenitors, and neurons. Several of these metabolites are normally present in human serum, suggesting their usefulness as potential serum FASD biomarkers. These findings suggest the biochemical pathways

that are affected by EtOH in the developing nervous system and delineate mechanisms of alcohol injury during human development.

A variety of possible mechanisms by which neurotoxicants can lead to neurodevelopmental abnormalities is supported by the scientific literature. These mechanisms involve induction of oxidative stress, interfering calcium signaling, effects on neurotransmitter pathways, neuroendocrine effects and epigenetic control (Chen et al., 2011). Especially during critical stages of nervous system development, the abovementioned effects can impact neuronal growth, differentiation, migration, synaptogenesis, and myelination, leading to an array of neurodevelopmental deficits. Many environmental toxicants, (heavy metals, PBDEs, PCBs, some pesticides etc.) possess the ability to generate ROS and deplete antioxidant capacity. Neuronal cells are especially vulnerable to oxidative stress because of the high amount of ROS generated during normal metabolism and neuronal activity (Milou et al. 2011). In vitro studies have demonstrated that environmental toxicant such as (Pb, Hg, PCBs, PBDEs) cause oxidative stress in neuronal cells leading to apoptotic cell death. Evidence is also present of the effect of many subclasses of pollutants to neurotransmitter pathways. Recently, an animal study discovered that repetitive postnatal PCB exposure resulted in increased levels of homovanillic and 5-hydroxyindoleacetic acid, metabolites of DA and 5-HT, in the neostriatum of young adult animals, without changing levels of the transmitters themselves. Moreover, analysis of DA-synapses demonstrated specific effects on a restricted number of specific synaptic proteins, including the presynaptic DAT, the postsynaptic D5 receptor and the PSD-95 scaffolding synapse protein(Dervola et al., 2015). Exposure to PCBs have been found to affect the levels of urinary homovanillic acid, a DA metabolite, in humans (Putschögl et al., 2015).

Autism Spectrum Disorders (ASD) are a group of developmental disorders caused by environmental and genetic factors. Diagnosis is based on behavioral and developmental signs detected before 3 years of age with no reliable biological marker. The potential use of a 2D NMR-based approach to

express the global biochemical signature of autistic individuals compared to normal controls was investigatedby Mavel et al. (2013). This technique has greater spectral resolution than to 1D H NMR spectroscopy, which is limited by overlapping signals. The urinary metabolic profiles of 30 autistic and 28 matched healthy children were obtained using a $^{1}H–^{13}C$ NMR-based approach. The data acquired were processed by multivariate orthogonal partial least-squares discriminant analysis (OPLS-DA). Some discriminating metabolites were identified: β-alanine, glycine, taurine and succinate concentrations were significantlyhigher, and creatine and 3-methylhistidine concentrations were lower in autistic children than in controls. Also, differences in several other metabolites that were not identified but characterized by a cross peak correlation in $^{1}H–^{13}C$ HSQC were noted. Statistical models of 1H and $^{1}H–^{13}C$ analyses were compared and only 2D spectra allowed the characterization of statistically relevant changes [R2Y(cum)¼ 0.78 and Q2(cum)¼ 0.60] in the low abundance metabolites. This method has the potential to contribute to the diagnosis of neurodevelopment disorders but needs to be validated on larger cohorts and on other developmental disorders to define its specificity. Similarly, Wang et al. (2016)performed a metabolomics analysis of serum to identify potential biomarkers for the early diagnosis and clinical evaluation of autism. They analyzed a discovery cohort of patients with autism and participants without autism in the Chinese Han population using ultra-performance liquid chromatography quadrupole time-of-flight tandem mass spectrometry (UPLC/Q-TOF MS/MS) to detect metabolic changes in serum associated with autism. The potential metabolite candidates for biomarkers were individually validated in an additional independent cohort of cases and controls. They built a multiple logistic regression model to evaluate the validated biomarkers, including 73 patients and 63 controls in the discovery cohort and 100 cases and 100 controls in the validation cohort. Metabolomics analysis of serum in the discovery stage identified 17 metabolites, 11 of which were validated in an independent cohort. A multiple logistic regression model built on the 11 validated metabolites fit well in both cohorts. The model consistently showed that autism was associated with 2 particular metabolites: sphingosine 1-

phosphate and docosahexaenoic acid. In another study (Emond et al., 2013)., GC-MS urinary metabolic profiles of 26 autistic and 24 healthy children were obtained by liq/liq extraction, and were or were not subjected to an oxidation step, and then were subjected to a persilylation step. These metabolic profiles were then processed by multivariate analysis, in particular orthogonal partial least-squares discriminant analysis (OPLS-DA, R2Y (cum) = 0.97, Q2 (cum) = 0.88). Discriminating metabolites were identified. The relative concentrations of the succinate and glycolate were higher for autistic than healthy children, whereas those of hippurate, 3-hydroxyphenylacetate, vanillylhydracrylate, 3-hydroxyhippurate, 4-hydroxyphenyl-2-hydroxyacetate, 1H-indole-3-acetate, phosphate, palmitate, stearate, and 3-methyladipate were lower. Eight other metabolites, which were not identified but characterized by a retention time plus a quantifier and its qualifier ion masses, were found to differ between the two groups. Comparison of statistical models leaded to the conclusion that the combination of data obtained from both derivatization techniques leads to the model best discriminating between autistic and healthy groups of children.

In terms of transcriptomics signatures, recent studies of genomic variation associated with autism have suggested the existence of extreme heterogeneity. Large-scale transcriptomics should complement these results to identify core molecular pathways underlying autism. Gupta et al. (2014) reported results from a large-scale RNA sequencing effort, utilizing region-matched autism and control brains to identify neuronal and microglial genes robustly dysregulated in autism cortical brain. A gene expression module corresponding to M2-activation states in microglia is negatively correlated with a differentially expressed neuronal module, implicating dysregulated microglial responses in concert with altered neuronal activity-dependent genes in autism brains. These observations provided pathways and candidate genes that highlight the interplay between innate immunity and neuronal activity in the etiology of autism.

In addition to metabolomics and transcriptomics signatures, Heberling and Dhurjati(2015) proposed a computational approach whereby metagenomes

characteristic of "healthy" and autistic individuals are artificially constructed via genomic information, analyzed for the enzymes coded within, and then these enzymes are compared in detail. This is a text mining application. A custom-designed online application was built and used for the comparative metabolomics study and made publically available. Several of the enzyme-catalyzing reactions involved with the amino acid glutamate were curiously missing from the "autism" microbiome and were coded within almost every organism included in the "control" microbiome. Interestingly, there exists a leading hypothesis regarding autism and glutamate involving a neurological excitation/inhibition imbalance; but the association with this study is unclear. The results included data on the transsulfuration and transmethylation pathways, involved with oxidative stress, also of importance to autism. The results from this study are in alignment with leading hypotheses in the field, which is impressive, considering the purely in silico nature of this study. The present study provides new insight into the complex metabolic interactions underlying autism, and this novel methodology has potential to be useful for developing new hypotheses. However, limitations include sparse genome data availability and conflicting literature experimental data. We believe our software tool and methodology has potential for having great utility as data become more available, comprehensive and reliable. In terms of computation methods, starting from plasma metabolome analysis from children aged 4 to 6, 52 with ASD and 30 age-matched TD children, a broad range of metabolites were identified (West et al., 2014). Univariate, multivariate and machine learning methods were used to develop models to rank the importance of features that could distinguish ASD from TD. A set of 179 statistically significant features resulting from univariate analysis were used for multivariate modeling. Subsets of these features properly classified the ASD and TD samples in the 61-sample training set with average accuracies of 84% and 86%, and with a maximum accuracy of 81% in an independent 21-sample validation set. In another study, a supervised multivariate model to classify the metabolome alterations between autistic spectrum disorders (ASD) patients and controls, siblings of autistic patients, had been realized and used to realize a network model of the ASD patients' metabolome(Noto et al., 2014).

In this experiment, quantification of urinary metabolites with Mass Spectroscopy coupled to Gas Chromatography was made. A multivariate model has been used to extrapolate the variables of importance for a network model of interaction between metabolites. In this way we are able to propose a network-based approach to ASD description. Plasma metabolomics phenotyping was obtained for 23 premutation carriers and 16 age- and sex-matched controls. Three biomarkers, phenylethylamine normalized by either aconitate or isocitrate and oleamide normalized by isocitrate, exhibited excellent model performance(Giulivi et al., 2016). The lower phenylethylamine and oleamide plasma levels in carriers may indicate, respectively, incipient nigrostriatal degeneration and higher incidence of substance abuse, anxiety and sleep disturbances. Higher levels of citrate, isocitrate, aconitate, and lactate may reflect deficits in both bioenergetics and neurotransmitter metabolism (Glu, GABA).

Biomarkers-Metabolites of neurodevelopmental disorders

Neurodevelopmental disorders with periventricular nodular heterotopia (PNH) are etiologically heterogeneous, and their genetic causes remain in many cases unknown. Missense mutations in NEDD4L mapping to the HECT domain of the encoded E3 ubiquitin ligase lead to PNH associated with toe syndactyly, cleft palate and neurodevelopmental delay(Broix et al., 2016). Cellular and expression data showed sensitivity of PNH-associated mutants to proteasome degradation. Moreover, an in utero electroporation approach showed that PNH-related mutants and excess wild-type NEDD4L affect neurogenesis, neuronal positioning and terminal translocation. Further investigations, including rapamycin-based experiments, found differential deregulation of pathways involved. Excess wild-type NEDD4L leads to disruption of Dab1 and mTORC1 pathways, while PNH-related mutations are associated with deregulation of mTORC1 and AKT activities. Altogether, these data provide insights into the critical role of NEDD4L in the regulation of mTOR pathways and their contributions in cortical development. De novo mutations in CHD8 are strongly associated with autism spectrum disorder, but the basic biology of CHD8 remains poorly understood. CHD8 knockdown

during cortical development results in defective neural progenitor proliferation and differentiation that ultimately manifests in abnormal neuronal morphology and behaviors in adult mice(Durak et al., 2016). Transcriptome analysis revealed that while CHD8 stimulates the transcription of cell cycle genes, it also precludes the induction of neural-specific genes by regulating the expression of PRC2 complex components. Furthermore, knockdown of CHD8 disrupts the expression of key transducers of Wnt signaling, and enhancing Wnt signaling rescues the transcriptional and behavioral deficits caused by CHD8 knockdown. These roles of CHD8 and the dynamics of CHD8 expression during development help negotiate the fine balance between neural progenitor proliferation and differentiation. Together, these observations provide new insights into the neurodevelopmental role of CHD8. Caubit et al. (2016)identified TSHZ3 as the critical region for a syndrome associated with heterozygous deletions at 19q12-q13.11, which includes autism spectrum disorder (ASD). In TSHZ3-null mice, differentially expressed genes include layer-specific markers of cerebral cortical projection neurons (CPNs), and the human orthologs of these genes are strongly associated with ASD. Furthermore, mice heterozygous for TSHZ3 show functional changes at synapses established by CPNs and exhibit core ASD-like behavioral abnormalities. These findings highlight essential roles for TSHZ3 in CPN development and function, whose alterations can account for ASD in the newly defined TSHZ3 deletion syndrome. A genome-wide microRNA (miRNA) expression profiling in post-mortem brains from individuals with ASD and controls and identified miRNAs and co-regulated modules that were perturbed in ASD was performed by Wu et al. (2016). Putative targets of these ASD-affected miRNAs were enriched for genes that have been implicated in ASD risk. A regulatory relationship between several miRNAs and their putative target mRNAs in primary human neural progenitors was confirmed. These include hsa-miR-21-3p, a miRNA of unknown CNS function that is upregulated in ASD and that targets neuronal genes downregulated in ASD, and hsa_can_1002-m, a previously unknown, primate-specific miRNA that is downregulated in ASD and that regulates the epidermal growth factor receptor and fibroblast growth factor receptor signaling pathways involved in neural

development and immune function. In a recent application of genome-wide association studies (GWAS) to ASD, Inoue and Inoue(2016)indicated significant associations with the single nucleotide polymorphisms (SNPs) on chromosome 5p14.1, located in a non-coding region between cadherin10 (CDH10) and cadherin9 (CDH9). An in vivo bacterial artificial chromosome (BAC) based enhancer-trapping strategy in mice to scan the gene desert for spatiotemporal cis-regulatory activities was applied. The results showed that the ASD-associated interval harbors the cortical area, striatum, and cerebellum specific enhancers for a long non-coding RNA, moesin pseudogene1 antisense (MSNP1AS) during the brain developing stages. Mouse moesin protein levels are not affected by exogenously expressed human antisense RNAs in transgenic brains, demonstrating the difficulty in modeling rather smaller effects of common variants. This in vivo evidence for the spatiotemporal transcription of MSNP1AS however provides a further support to connect this intergenic variant with the ASD susceptibility. With regard to human studies, to identify candidate genes for intellectual disability a meta-analysis on 2,637 de novo mutations, identified from the exomes of 2,104 patient–parent trios was performed (Lelieveld et al., 2016). Statistical analyses identified 10 new candidate ID genes: DLG4, PPM1D, RAC1, SMAD6, SON, SOX5, SYNCRIP, TCF20, TLK2 and TRIP12. In addition, it was showed that these genes are intolerant to non-synonymous variation and that mutations in these genes are associated with specific clinical ID phenotypes. In addition, mutations in the aspartate/glutamate mitochondrial transporter, SLC25A12, have been associated with ASD (West et al., 2014).

Attention-deficit hyperactivity disorder (ADHD) is a prevalent and highly heritable disorder of childhood with negative lifetime outcomes. Although candidate gene and genome-wide association studies have identified promising common variant signals, these explain only a fraction of the heritability of ADHD. The observation that rare structural variants confer substantial risk to psychiatric disorders suggests that rare variants might explain a portion of the missing heritability for ADHD. A large-scale next-generation targeted sequencing study of ADHD in 152 child and adolescent

cases and 188 controls across an a priori set of 117 genes was performed by Hawi et al. (2016). A multi-marker gene-level analysis of rare (<1% frequency) single-nucleotide variants (SNVs) revealed that the gene encoding brain-derived neurotrophic factor (BDNF) was associated with ADHD at Bonferroni corrected levels. Sanger sequencing confirmed the existence of all novel rare BDNF variants. BDNF is a genetic risk factor for ADHD, potentially by virtue of its critical role in neurodevelopment and synaptic plasticity.

Bioinformatics

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data, combining computer science, statistics, mathematics and engineering to analyze and interpret biological data. It has been used for in silico analyses of biological raw datathrough mathematical and statistical techniques.

Bioinformatics is important for biological studies. In most cases, bioinformatics uses include the identification of candidate genes and nucleotides (SNPs), the gene and protein expression, the analysis of cellular organization, the structural and interaction network analysis. The aim of bioinformatics is the better understanding of genetic basis of diseases, unique adaptations, desirable properties (esp. in agricultural species), differences between populations and design methods and drugs for improvement of health. Bioinformatics includes large amounts of data (sequence, structure, function, biochemical information, protein-protein interactions, protein-DNA complexes, kinetics of reactions) which integrated together into Systems Biology, allow us to study the interactions between the components of a biological system and how those interactions give rise to the function and behavior that we can see.

A biological database is a large, organized body of persistent data, usually associated with computerized software designed to update, query, and retrieve components of the data stored within the system. A simple database might be a single file containing many records, each of which includes the same set of information. The first bioinformatics /biological databases were

constructed a few years after the first protein sequences began to become available. The first protein sequence reported was that of bovine insulin in 1956, consisting of 51 residues. After the formation of the databases, tools became available to search sequence databases - at first in a very simple way, looking for keyword matches and short sequence words, and then more sophisticated pattern matching and alignment based methods. Currently, a lot of bioinformatics work is concerned with the technology of databases. These databases include both "public" repositories of gene data like GenBank or the Protein DataBank (the PDB), and private databases like those used by research groups involved in gene mapping projects or those held by biotech companies. Bioinformatic tools are software programs that are designed for extracting the meaningful information from the mass of data & to carry out this analysis step.

Bioinformatics also tries to fingerprint the role of metabolites in human body, creating a large number of databases. (Figure 2)

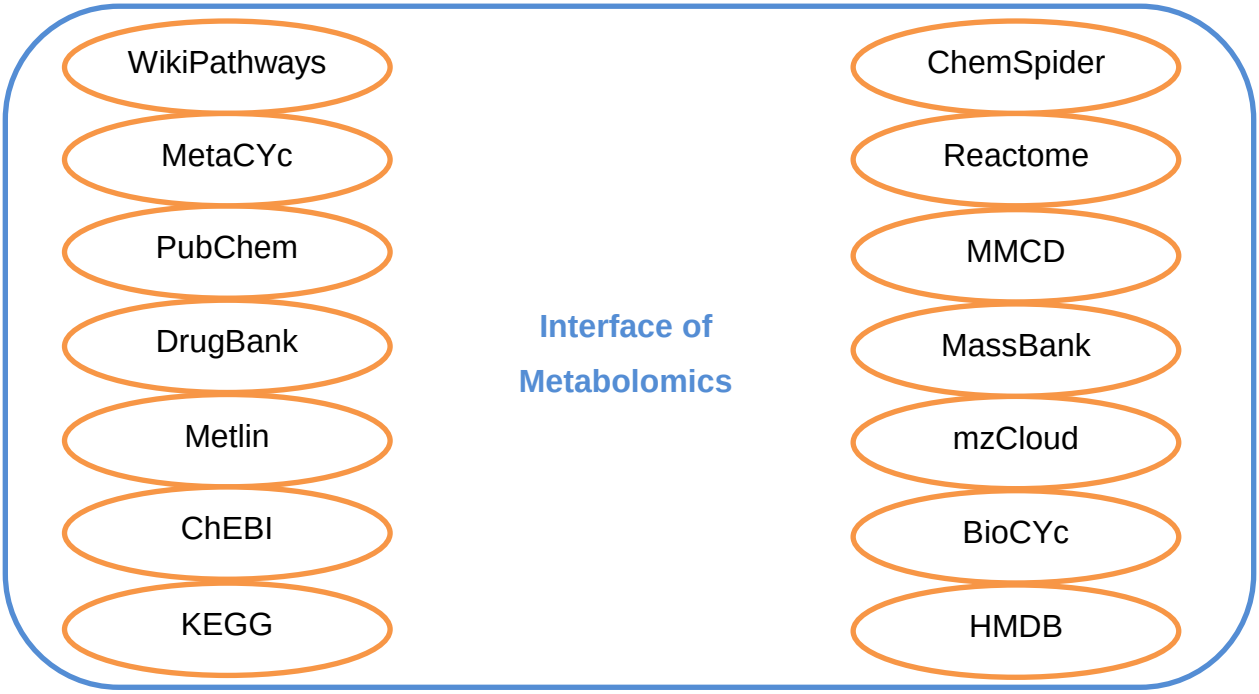


Figure 2: Metabolomics Databases

Environmental health associations

Environmental health has been defined in a 1999 document by the World Health Organization (WHO) as: Those aspects of the human health and

disease that are determined by factors in the environment. It also refers to the theory and practice of assessing and controlling factors in the environment that can potentially and affect health. Environmental health as used by the WHO Regional Office for Europe, includes both the direct pathological effects of chemicals, radiation and some biological agents, and the effects (often indirect) on health and well being of the broad physical, psychological, social and cultural environment, which includes housing, urban development, land use and transport(Novice, 1999).

As of 2016 the WHO website on environmental health states "Environmental health addresses all the physical, chemical, and biological factors external to a person, and all the related factors impacting behaviors. It encompasses the assessment and control of those environmental factors that can potentially affect health. It is targeted towards preventing disease and creating health-supportive environments. This definition excludes behavior not related to environment, as well as behavior related to the social and cultural environment, as well as genetics." (WHO, 2015)

Three basic fields contribute to the field of environmental health: environmental epidemiology, toxicology, and exposure science. Each of these contributes different information to describe problems in environmental health.

Environmental epidemiology studies the relationship between environmental exposures (including exposure to chemicals, radiation, microbiological agents, etc.) and human health. Observational studies, which simply observe exposures that people have already experienced, are common in environmental epidemiology because humans cannot ethically be exposed to agents that are known or suspected to cause disease (Eyles, 2016).

Toxicology studies how environmental exposures lead to specific health outcomes, generally in animals, as a means to understand possible health outcomes in humans. Toxicology has the advantage of being able to conduct randomized controlled trials and other experimental studies because they can

use animal subjects(Eyles, 2016). However there are many differences in animal and human biology, and there can be a lot of uncertainty when interpreting the results of animal studies for their implications for human health.

Exposure science studies human exposure to environmental contaminants by both identifying and quantifying exposures. Exposure science can be used to support environmental epidemiology by better describing environmental exposures that may lead to a particular health outcome, identify common exposures whose health outcomes may be better understood through a toxicology study, or can be used in a risk assessment to determine whether current levels of exposure might exceed recommended levels(Eyles, 2016).

Information from these three disciplines can be combined to conduct a risk assessment for specific chemicals, mixtures of chemicals or other risk factors to determine whether an exposure poses significant risk to human health (exposure would likely result in the development of pollution-related diseases). This can in turn be used to develop and implement environmental health policy that, for example, regulates chemical emissions, or imposes standards for proper sanitation(Frumkin, 2010).

Why from Taranto, Italy?

Taranto is a coastal city in Apulia, Southern Italy and one of the areas identified at high risk of environmental crisis in Italy because of a wide industrial area developed nearby the urban settlement, with a high population density. The industrial zone is characterized by the presence of various types of plants, namely: one of the greatest steel mills in Europe, a major oil refinery, shipbuilding, a navy arsenal, a cement plant, two thermoelectric power plants and other plants for the manufacturing of rubber and plastic products, industrial chemicals, miscellaneous products of petroleum and coal, metal products, electric and electronic machinery and equipment (Marinaccio et al., 2011). Such industrial activities are responsible for environmental contamination, mainly due to heavy metals, asbestos, polycyclic aromatic

hydrocarbons (PAHs), organic solvents, polychlorinated biphenyls (PCBs) and dioxin (Marinaccio et al., 2011).

Studies of assessment of air emissions have been carried out in Taranto, in particular for suspended particulate matter in the neighborhood of steelworks, cement production and refinery plants, and for PAHs, benzo(a)pyrene, heavy metals released from the steel foundry's coke-oven batteries and from the municipal solid waste incinerator (Viviano et al., 2005). Occupational exposure to PAHs and inorganic arsenic has been investigated through biological monitoring on a sample of workers(Bisceglia et al., 2005).

For these reasons, it is important to notice how and in which grade an environment like this affects the emergence of neurodevelopmental disorders.

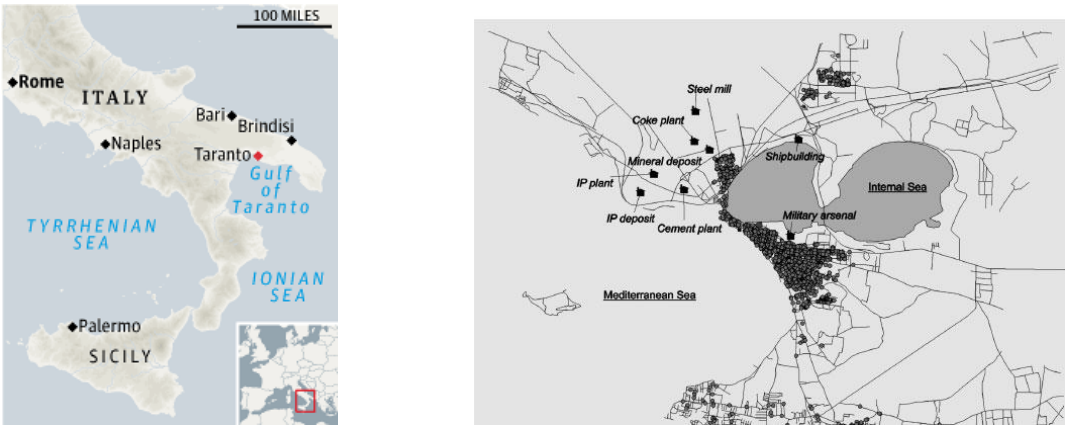
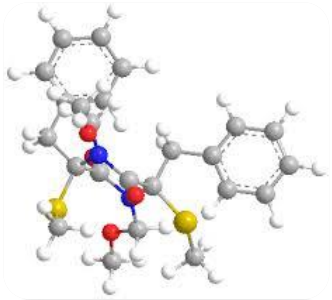
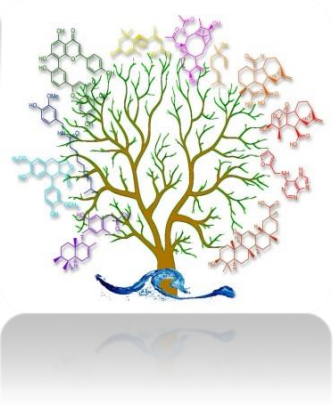


Figure3: **A.** The position of Taranto in Italy (left)
B. Distribution and distance zone from industrial sites in Taranto area (right)

Part II

Approach





Aim - Objective

To review recent evidence relating environmental exposures, through specific omics protocols involved the optimization of sample preparation requirements (untargeted metabolomics) to neurodevelopmental disorders. This also provides`t1 an overview for clinicians interested to understand the contribution of genetic heterogeneity in relation to environmental exposure and health outcome.

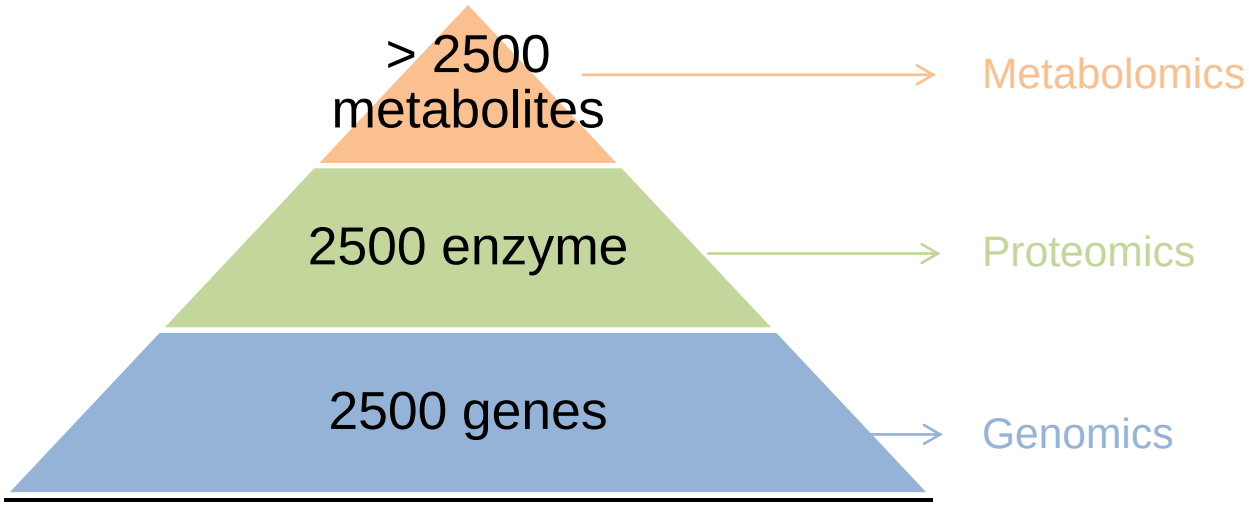


Figure 4: The pyramid of life by KEGG Database.

Data Sources

Pubmed(NCBI)was searched using the term *metabolomics* combined with the keywords *neurodevelopmental disorders, LC-MS/MS method*. Articles were selected based on relevance to the goals of this project. Studies that involved humans were prioritized, including routes and levels of exposure, developmental and early-life exposures, immunotoxicity, and the development of neurotrophic disorders. Moreover for the identification of each metabolite and the metabolic pathway were used METLIN Metabolite Database, Human Metabolome Database (HMDB), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Figure 4), BioCyc, PubChem and ChemSpider Database.

Materials

Sample Collection

For best results, the dietary diversity in the human population must be minimized. All subjects should fast overnight or refrain from food before collection of urine or blood samples. A brief description of consumed food should be included in any report. One approach that is currently used in clinical metabolomics studies is to make all subjects fast overnight and void the bladder first thing in the morning. The subject should then drink 500mL of water and avoid any other fluid or foodstuff for at least an hour, before voiding the bladder once again.

This is the metabolically 'clean' sample; assuming that we know there are starvation components (ketone bodies, etc.) that begin to appear. The presence of starvation components increases greatly if starved for an additional 4 hours. For clinical studies that involve patients with severe disease, it may be considered undue burden, or unhealthy, to request dietary restriction for minimization of potential diet-related influences on metabolomics profiles. Another approach is to recruit controls based on life-style factors following the collection of such information for the patient population.

For all clinical studies, independent of dietary restriction, it is desirable to recruit for control populations that provide best controls based on gender, age, and ethnic origin. For effective metabolomics analysis it is crucial that all samples are handled properly in order not to compromise the quality and accuracy of data acquired.

In this study, we investigated 92 urine samples from children. The ages of the human subjects ranged from 6 to 12 years, which were born and live in Taranto, a coastal city in Apulia, Southern Italy. Half of the child population were males. The average age of mother at birth is 33 years; Socioeconomic Status (SES) correlates with maternal and paternal education, breast feeding index and stress events index. Researchers indicate that SES is a key factor

that influences quality of life for children and families. SES affects human functioning in many ways, including development across the life span, psychological health, and physical health.

Methods

Chemicals and Reagents

Formic acid, ultrapure water and methanol were purchased from Sigma–Aldrich Company. All solvents used were LC–MS grade. Reserpine, caffeine and water were from LC Standards as their ion chromatograms are known.

Sample Preparation, Urine

If possible at least 1 mL of urine should have taken into a sterile container, aliquoted into 3 x Eppendorf tubes and immediately frozen before transportation. Volume and time of sampling should have documented. If possible at sampling site, the addition of 1 % sodium azide should have undertaken in at least one of the urine aliquots. The samples were kept after collection as cold as possible e.g. on ice, freezed as quickly as possible, preferably in liquid nitrogen and stored samples at -80 ^{o}C at least for metabolomic analysis.

Once in the laboratory the urine sample was thawed immediately prior to analysis before centrifugation at 1000 rpm for 10 minutes at room temperature. Supernatant (250 μL) was mixed with 500 μL of chilled water and analysed by Liquid chromatography – High Resolution Mass Spectrometry (LC-HRMS). (Figure 5)

UPLC/MS-MS analysis

The success in each experiment is the criteria in which the scientist selects the applicable methods developed in the laboratory. The triptych appropriate selectivity, suitable sensitivity and throughput speed is the baseline of the effectiveness in all scientific experiments. In more detail, a method should detect the substance of interest and do not react with other substances and the sensitivity must be ten times less than the parametric value. Last but not

least, a significant part is the speed from the sampling time until outcome effect shall, which is important for routine analyzes.

Ultra performance liquid chromatography – tandem mass spectrometry (UPLC-MS-MS) is an hyphenated analytical technique that combines the separation capabilities of liquid chromatography with the sensitivity and selectivity of mass spectrometry.

Ultra performance liquid chromatography utilizes very small particles packed and operating at relatively high pressure. Significant advances in chromatography technology were made to achieve dramatic increases in resolution, speed and sensitivity in liquid chromatography. In UPLC, the sample is forced by a liquid at high pressure (the mobile phase) through a column that is packed with a stationary phase generally composed of irregularly or spherically shaped particles chosen or chemically modified to accomplish particular types of separations. UPLC methods are divided into two different sub-classes based on stationary phases and the corresponding required polarity of the mobile phase – NP and RP.

Mass spectrometry (MS) is an analytical technique that measures the mass-to-charge ratio of charged particles(Sparkman, 2000). MS works by ionizing chemical compounds to generate charged molecules or molecule fragments and measuring their mass-to-charge ratios. The components of the sample are ionized by electrospray ionization (ESI), which results in the formation of charged particles (ions) in gas phase. The ions are separated according to their mass-to-charge ratio in an analyzer by electromagnetic fields. These include identifying unknown compounds, determining the isotopic composition of elements in a molecule, and determining the structure of a compound by observing its fragmentation.

LC conditions

The LC column used is an ACE 3Q 150 x 3 mm, 3 μm (Advanced Chromatography Technologies, Aberdeen, UK). Mobile phases are 0.1%

formic acid in LC-MS H_2O (mobile phase A, MPA) and 0.1% formic acid in MeOH (mobile phase B, MPB). Formic was used to reduce the pH of the mobile phase, to suppress the ionization of the weak organic acids, and to improve retention. Gradient applied is 99% MPA for 1 minute before increasing to 95% MPB over 9 minutes (Table 1). This is held for 1minutesbefore reverting back to 99% MPA and held for 4 minutes. The choice of the column is dependent on the compounds being analyzed. In urine samples the components are predominantly of low molecular mass and hydrophilic. So we have used: Acquality UPLC HSS T3 1.8μm, 2.1x100mm column and Acquality UPLC HSS T3 1.8μm VanGuard Pre-Column, 2.1x5mm. Injection volume is 5 μl, flow rate is 500 μl/min, column and sample temperature set to 40°C and 4°C respectively.

Table 1: Gradient elution program used for the LC separation, running in LC.

Time (min)	Mobile phase A	Mobile phaseB	Flow rate, μl/min
0	99	1	500
1	99	1	500
3	85	15	500
6	50	50	500
9	5	95	500
10	5	95	500
10.1	99	1	500
14	99	1	500

MS conditions

The MS used is a LTQ-Orbitrap Discovery (Thermo Fisher Scientific, MA, USA) set at 30,000 resolution and full width at half maximum (FWHM) 100 - 1000 m/z and high resolution and aquarate mass at 5ppm. The MS set up had the following options: Seath gas flow rate: 40, Aux. Gas flow rate: 8,Capillary temperature: 320^{o}C, Tube lens:120 V (positive ionization) / -120 V (negative ionization). Analysis will be undertaken in both positive and negative ionisation

mode (separate experiments). Every 70 runs, we paused the instrument mode for electrospray ionization (ESI)source cleaning and calibration.

Analytical procedure

One hour before starting each sample analysis, it is necessary to perform calibration of the instrument using a commercial reagent kit, the Proteo Mass LTQ/FT-Hybric ESI Neg. or Pos. Mode Cal Mix (Superlco, USA), dependent of the negative or positive instrument mode used run. One way to randomize the samples is to use a randomized block design, constructed to reduce noise or variance in the data(Want et al., 2010). The samples are divided into subgroups with 10 samples. It is good practice to run some pooled 'quality control' (QC) samples at the beginning of the run and use the data derived from these samples to demonstrate the system suitability, seeing that QC samples should behave in exactly the same way as the test. Moreover, it is necessary to inject known samples to run controlling system suitability and system stability. In the frame of this study, these associations were caffeine and reserpine with known chromatography pattern. At the end of each run the column was washed thoroughly with a strongly eluotropic solvent (methanol),and the MS inlet and source meticulously cleaned before the next run to prevent the build up of contaminants and ensure continuing good performance(Want et al., 2010). A typical QC sample, would be a pooled urine sample, prepared by mixing aliquots of the samples to be analyzed and therefore broadly representative of the whole sample set (Nordstrom et al., 2008). A typical sample sequence for both positive and negative runs was:

Sample Reserpine 1ppm

Sample Caffeine 20ppm

Blank Water

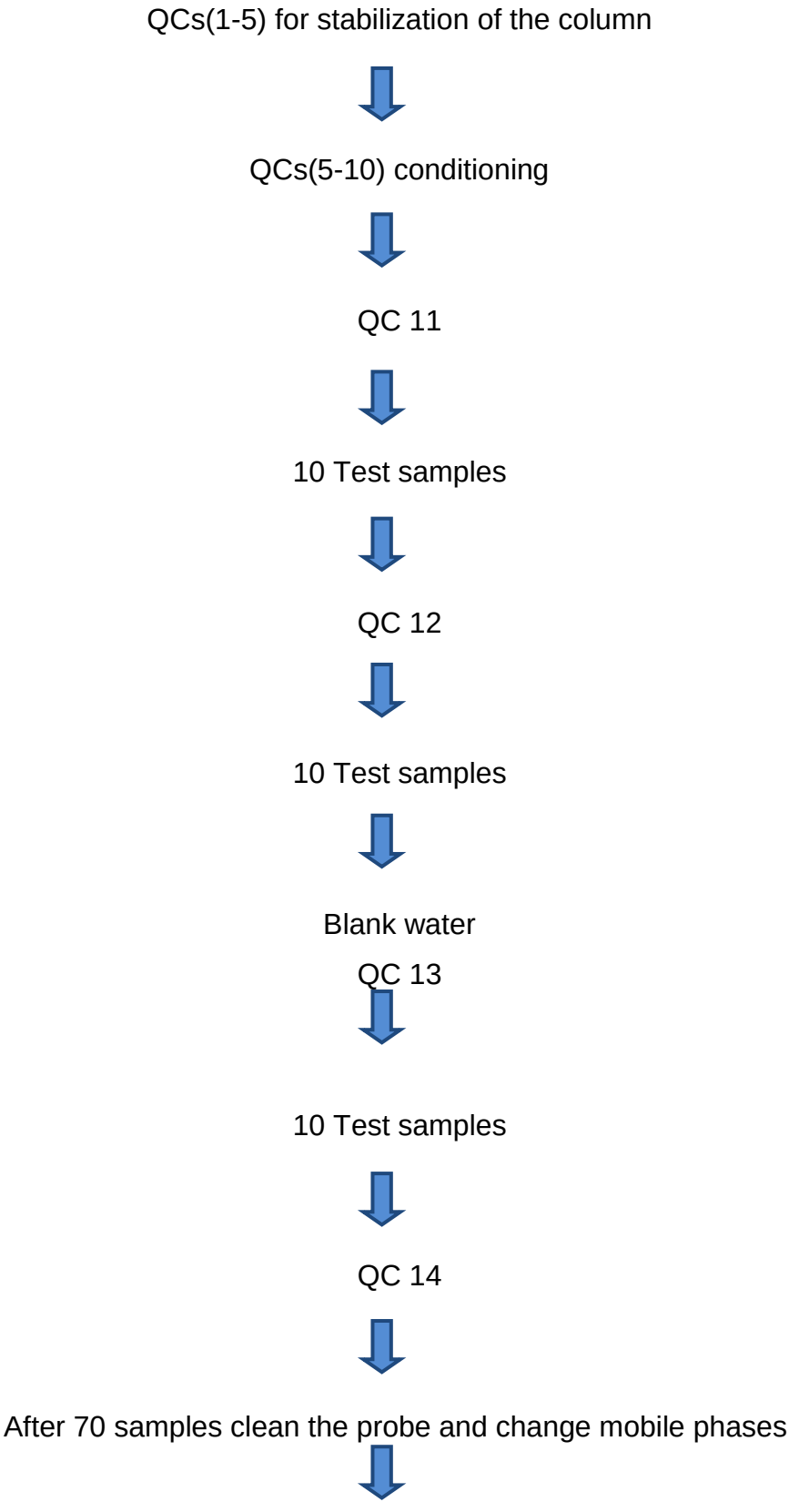


Figure 5: Samples sequence applied for the analysis

Totally, 127 runs analyzed both positive and negative ionization.

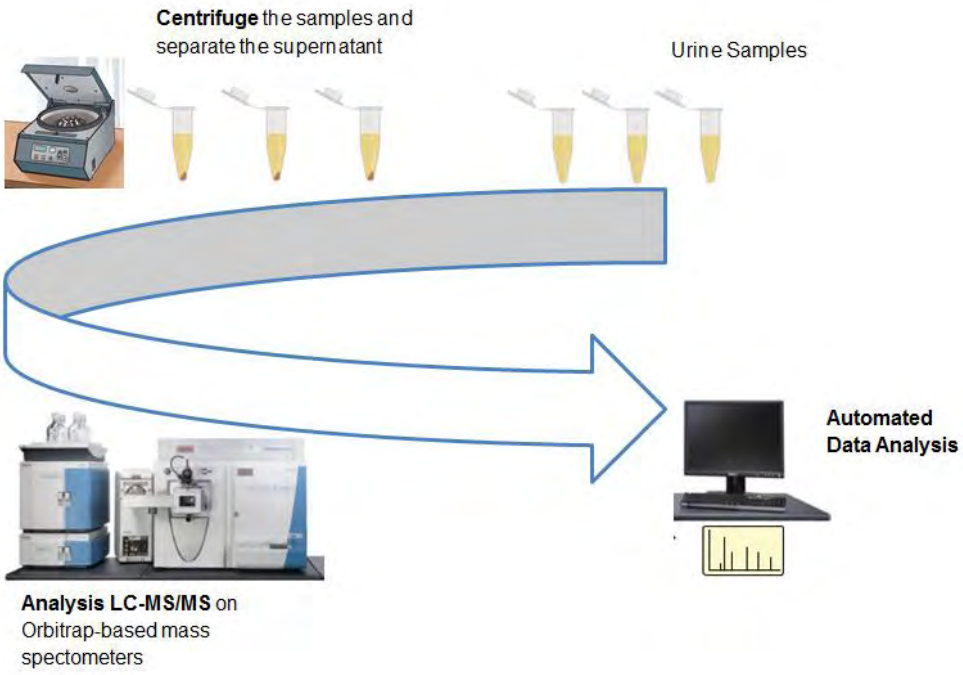


Figure6: Untargeted metabolomics workflow

Analysis output

Xcalibur 2.1 software (ThermoFisher Scientific Inc., San Jose, CA, USA) was used to analyze and process all data for quantitative analysis.This software package also offers data reduction techniques to de-isotope / de-adduct features selected in all data files before the option of sample differentiation by multivariate and univariate statistical analysis. Where necessary, data

Figure7: A. The LC-MS/MS system in Environmental Engineering Laboratory (AUTH) (left)
B. Cooled sample space (right)

features (includes m/z's, retention times and peak intensity, often peak area) were normalized (using a script in R) using the QC data to correct for any within or between batch variation.

Once a satisfactory list of features are produced, tentative identification based on accurate mass only was applied using in-house or online databases (e.g. Human Metabolome Database).

Data Analysis

MZminesoftware

MZmine was developed at Okinawa Institute of Science and Technology, Japan and VTT Finland(Katajamaa et al., 2006b). It is a Java based program and is therefore platform independent. MZmine is an open-source framework for processing, visualization and analysis of mass spectrometry based molecular profile data.

This tool is available for manipulating the raw data files, including methods for noise reduction by filtering in chromatographic direction, cropping raw data range and removing scans by their width. The stages of spectral data processing are sequential. Following peak detection, many of the peaks have none or only few matches in other samples. The reasons for the misses vary. The peak may not be present in the sample, the peak detection may have failed because of noisy raw data or inaccurate parameter settings may have been used for peak detection and chromatographic alignment methods. This method first searches for a local intensity maximum within a selected chromatographic region corresponding to expected location of a missed peak, which is used as an estimate for peak height(Katajamaa et al., 2006a). The peak area estimate is then calculated by moving from the maximum to both directions along the extracted ion chromatogram as long as the peak curve is monotonously decreasing within the pre-specified tolerance limits(Katajamaa et al., 2006a).

Metabololites Search

Mass-based search is an important step for metabolite identification in mass-spectrometry-based metabolomic analysis. The mass-to-charge ratio (m/z) value of a molecular ion of interest is searched against metabolite database(s). The metabolites having molecular weights within a specified tolerance to the query m/z value are retrieved from the databases as putative identifications. These putative identifications serve as a foundation for further metabolite verification. In addition to searching with m/z values only, the ion annotation information can be used to aid the mass-based search. Ion annotation groups the ions originating from the same metabolite together and annotates them as adducts/isotopes/in-source fragments.(Zhou et al., 2012)

In this study we used MetaboSearch (Zhou et al., 2012)to perform mass-based metabolite search simultaneously against the four major metabolite databases: Human Metabolome DataBase (HMDB), Madison Metabolomics Consortium Database (MMCD), Metlin, and LipidMaps. The search results from these databases are integrated into a uniformly and non-redundant format based on IUPAC International Chemical Identifier (InChI) key. Chemical identifier information is provided by the software for effective reference to metabolites. Cross-referencing across multiple databases is performed when a particular identifier type is missing from a database. The comprehensive list of chemical identifiers includes PubChem Compound ID (CID), PubChem Substance ID (SID), HMDB ID, KEGG ID, InChI string, and InChI key. MetaboSearch performs mass-based search using a given list of m/z values.(Zhou et al., 2012)

In the MetaboSearch interface, the top panel shows the four databases used for searching (HMDB, Metlin, MMCD, LipidMaps). The user has the option to include or exclude any particular database from the search. In this study, was checked the four available databases in which the tool would search about the metabolites. Middle left panel is the input area for data uploading and searching parameters. Set the ionization mode of the data by checking either 'positive' or 'negative' when the file with the negative or positive ionization was

added respectively. Our last selection was the m/z tolerance in 10 ppm. After the submission of the file and the program's running, in output frame were displayed the intermediate results. The bottom panel displays the running status and the progress. In this way were identified a large number of metabolites.

After the identification, we needed to find the CasID of each metabolite from Human Metabolite Database (HMDB). CasID is a specific and unique number for each metabolite.

MassProfilerPro

Agilent Mass Profiler Professional (MPP) software is a powerful chemometrics platform designed to exploit the high information content of mass spectra (MS) data and can be used in any MS-based differential analysis to determine relationships among two or more sample groups and variables. MPP provides advanced visualization tools for GC/MS, LC/MS,CE/MS, ICP-MS, and NMR data analysis. MPP also integrates smoothly withAgilent MassHunter Workstation, Spectrum Mill and ChemStation software and is the only platform that provides integrated identification/ annotation of compounds and integrated pathway analysis for metabolomic and proteomic studies. The system also enables.

Creating a new project, the user can import the data both from positive and negative ionization and the necessary parameters. Setting the minimum absolute tolerance (30 counts), the alignment parameters, the normalization and the baseline of the experiment the data are grouped. The next step is the pathway analysis, selecting the suitable databases. At the end, all the pathways in which the studied metabolites are implicated, appear in the MPP interface.

Statistical Analysis-EWAS

Genetic expression coupled with environmental and epigenetic influences result in the phenotype expression of each person. Thus, it is important to link environmental exposures, the genome and consequently the biological pathways induced with the origin or evolution of disease. These factors play a key role in human health status. Along this line, Wild (2005)proposed a new way to conceptualize the environment called the “exposome”, consisting of the ensemble of exposures from birth to death. Recently, Patel and his team at Harvard University have crafted a new method to search for environmental factors associated with disease called the environment-wide association study (EWAS), which is analytically analogous to the genome-wide association study (GWAS) and also a comprehensive way to search for genetic variants associated with disease (Patel and Manrai, 2015).

The heterogeneous of environmental exposures depends from time, geographic location, climate and human impact on ecosystem, which influence the genetic variants in organisms. Importantly, given an exposure identified from an EWAS, it is very difficult to infer if the exposure is independently associated with the disease, the direction of association(“what causes what”), or if the exposure is simply a correlate (Ioannidis et al., 2009; Patel and Ioannidis, 2014). The aim of EWAS was to develop a method to identify robust correlations between exposures which leads that the correlation between exposures may allow investigators to describe how exposures can lead to other exposures (Patel and Manrai, 2015) and to describe correlations between exposure variables to construct an “exposome globe”, extending methods developed for unsupervised learning with genomic data called “relevance networks”(Butte and Kohane, 2000).

This study reveals the correlation between metabolic pathways which are searched through detected metabolites in urine samples and environmental exposures, computed the non-parametric correlation coefficient between each pair of factors (Patel and Manrai, 2015).We used a permutation-based approach to estimate the two-sided p-value of significance for each pair of

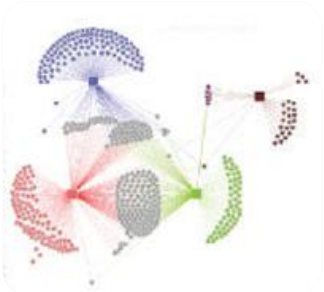
correlations. Specifically, each environmental factor was randomly permuted (sampled without replacement) and the correlations were re-computed to create a set of correlations that reflected the null distribution of no correlation.

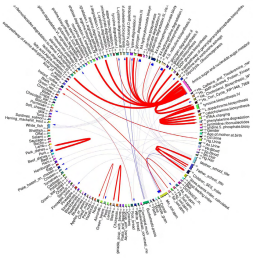
In the framework of this study, a variation of the original EWAS method of Patel and Manrai (2015) developed by Prof. Sarigiannis' lab was applied in this thesis on neurodevelopmental disorders. We searched for associations between 74 different biological pathways with 89 environmental, social and nutritional factors. Neurodevelopmental progress was evaluated using the Wechsler Intelligence Scale for children – Fourth Edition(Wechsler, 2003). This is an individually administered measure of intelligence intended for children aged six years to 16 years and 11 months. WISC-IV yields measures of general intelligence as reflected in both verbal and nonverbal (performance) abilities and specific indices including verbal comprehension, perceptual reasoning, working memory and processing speed.

We visualized the EWAS findings from these studies in the exposome globe. First, we plotted the –log10(p-value) of association between the factor and outcome (e.g. pathways) as a scatter plot in the R Circos(Zhang et al., 2013) plot (referred to as an "EWAS track" below).

Part III

Outcome

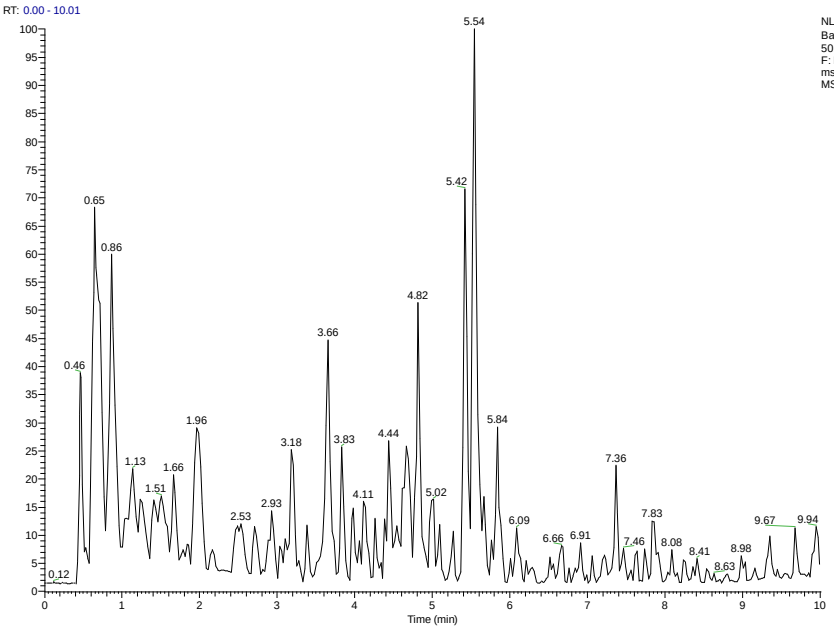




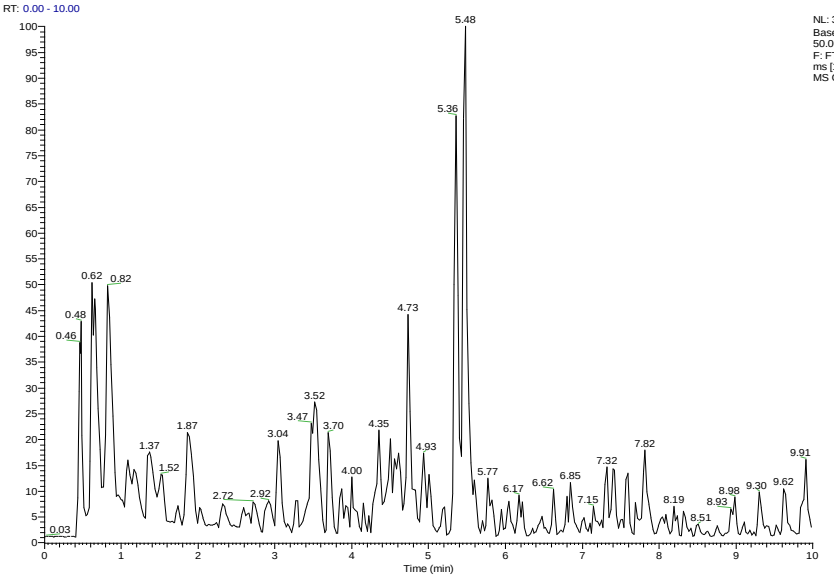
Results

LC/MS-MS

After LC-MS/MS analysis, we had the chromatogram which is characteristic fragmentation pattern for each sample. Typical chromatograms in positive ionization and negative ionization in samples are depicted in figure 7A, 7C. These chromatograms constitute the result of LC-MS/MS. Figures 7B, 7D present the positive and negative ionization in QC samples.

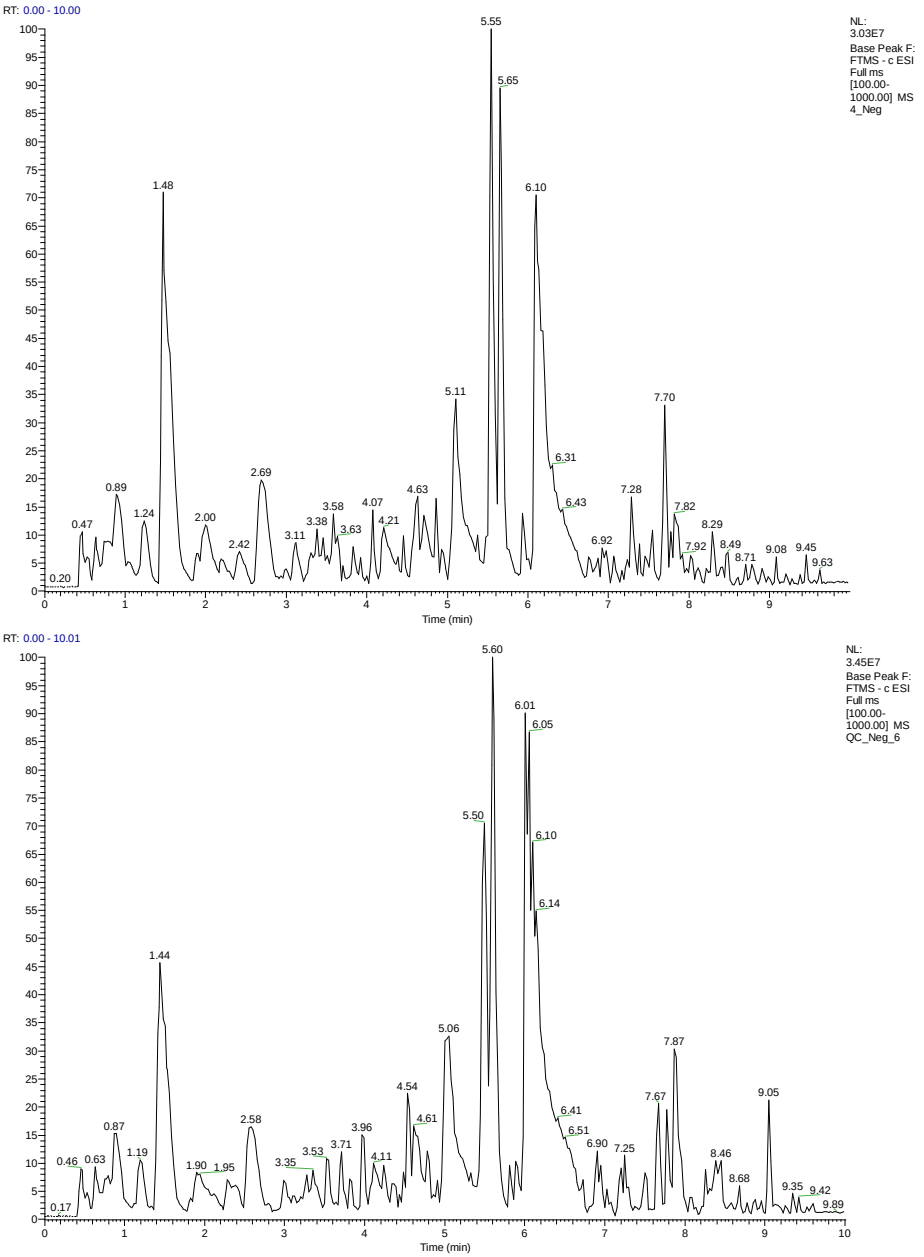


(A)



(B)

(B)



(C)

(D)

Figure 8: The chromatograms for a sample on A and a QC on B for positive mode and for a sample on C and a QC on D for negative mode respectively.

MZmine

Using MZmine on the output of LC-MS/MS we identified the metabolites detected in each sample. The output of MZmine was an excel file in which for each sample the m/z ratio, retention time, ID in HMDB, Identification method, Molecular Formula, Name and the number of the raw bc peak area were given.

Metabosearch

In the next stage, the name and the formula of each metabolite are confirmed through Metabosearch. Search of CasID in the Human Metabolite Database follows.

Mass Profiler Pro

Using the output of the previous bioinformatics steps, MPP is linked with on-line databases for searching biological pathways in which the identified metabolites are involved. A typical interface of this program is shown in figure 8. On the left side there is a list of pathways which were involved in each sample and in the right is shown analytically the route of each pathway.

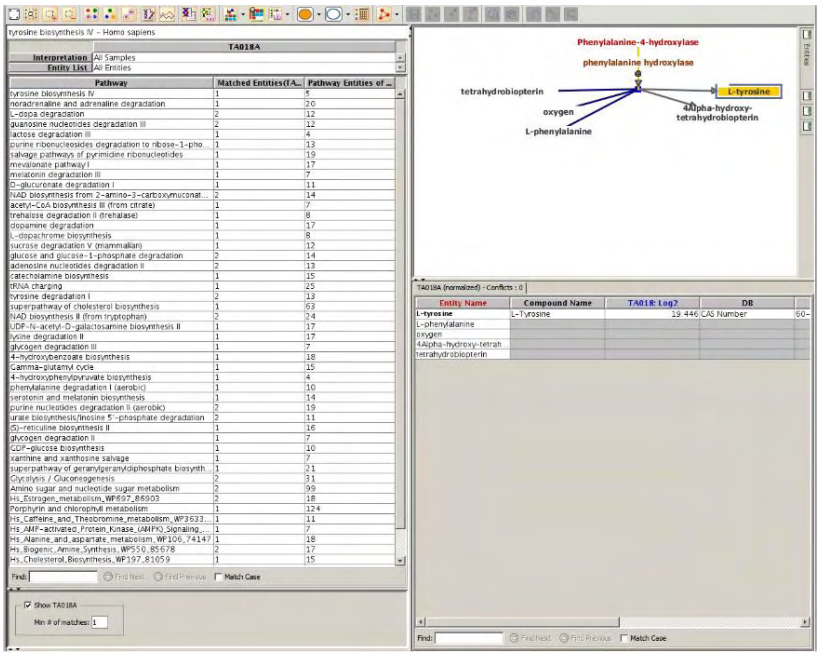


Figure 9: The output of MPP.

EWAS

Environment-wide association studies (EWAS) provide a way to uncover the environmental mechanisms involved in complex traits in a high-throughput manner (Hall et al., 2014). We systematically and comprehensively assessed the association of 89 unique factors (environmental, social and nutritional) measured in the Taranto cohort with 74 specific metabolic pathways which are related to neurodevelopmental disorders.

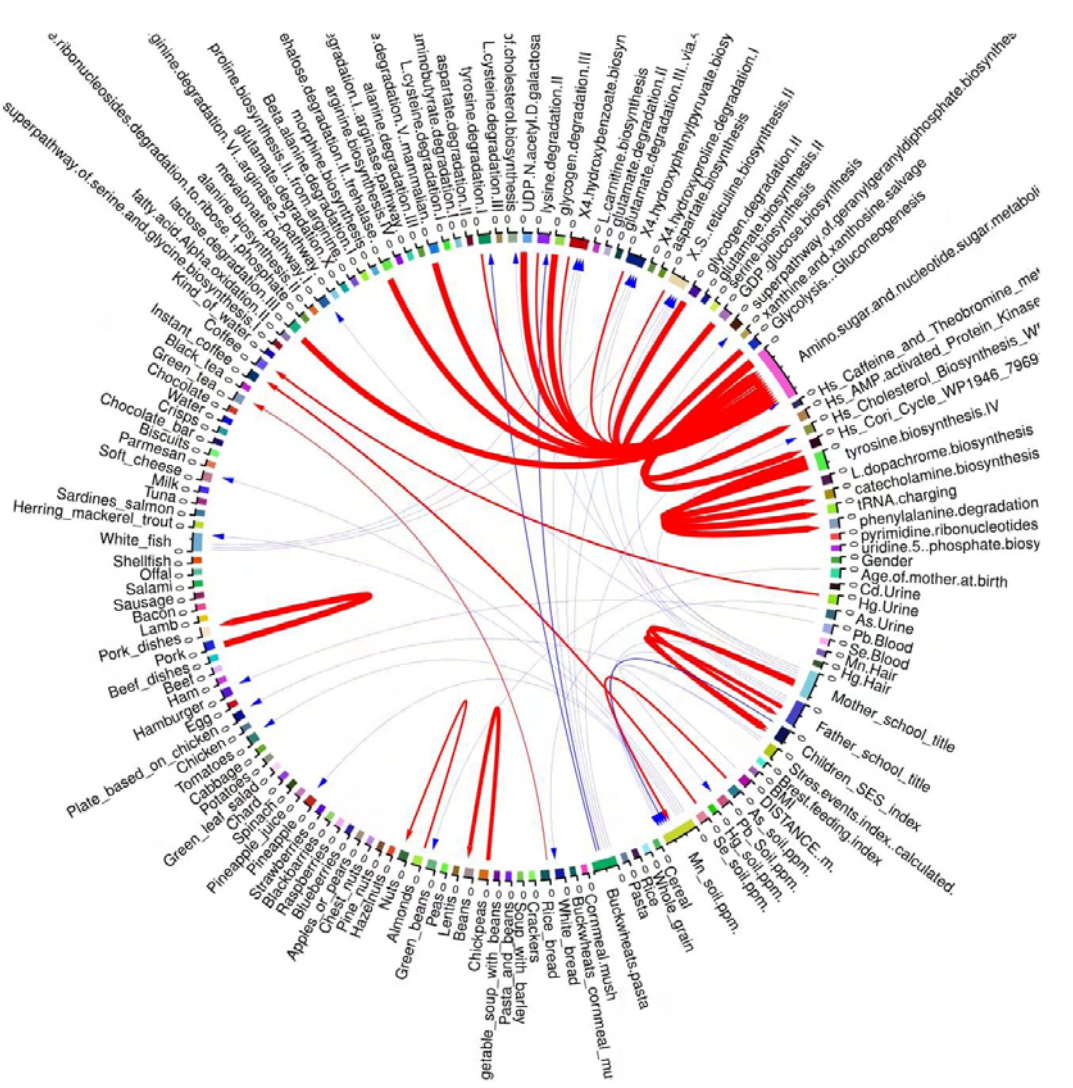


Figure 10: Correlation Globes for EWAS in Neurodevelopmental disorders. 89 different factors and 74 pathways are depicted in different colors in the globe. Association p-values from EWAS are shown as a separate track above each exposure (red points denote EWAS validated associations with positive effect size [indicating risk] blue points indicate an EWAS validated negative effect size [indicating protective]). Line thickness is proportional to size of the absolute value of correlation coefficient.

We observe interpretable broad patterns in the correlation globe (Figure 10). In the globe, strong correlations are depicted among several pathways, with Amino Sugar and Nucleotide Sugar Metabolism appearing to connect with several other pathways (red correlation lines). Moreover, we observe that the biosynthesis of tyrosine is also positively associated with other four pathways,

the biosynthesis of L-dopachrome and catecholamine, the degradation of phenylalanine and the tRNA charging (red correlation lines). On the other hand, negative correlations are observed between consumption of buckwheat pasta and tyrosine degradation, lysine degradation III and the pathways of 4-hydroxybenzoate biosynthesis, 4-hydroxyphenylpyruvate biosynthesis and (S)-reticuline biosynthesis II (blue correlation line). White fish in diet was negatively correlated with the pathways of 4-hydroxybenzoate biosynthesis, 4-hydroxyphenylpyruvate biosynthesis and (S)-reticuline biosynthesis II (blue correlation line). There is also a correlation between demographic parameters and metabolic pathways. Briefly, mother and father's educational level were correlated negatively with some pathways.

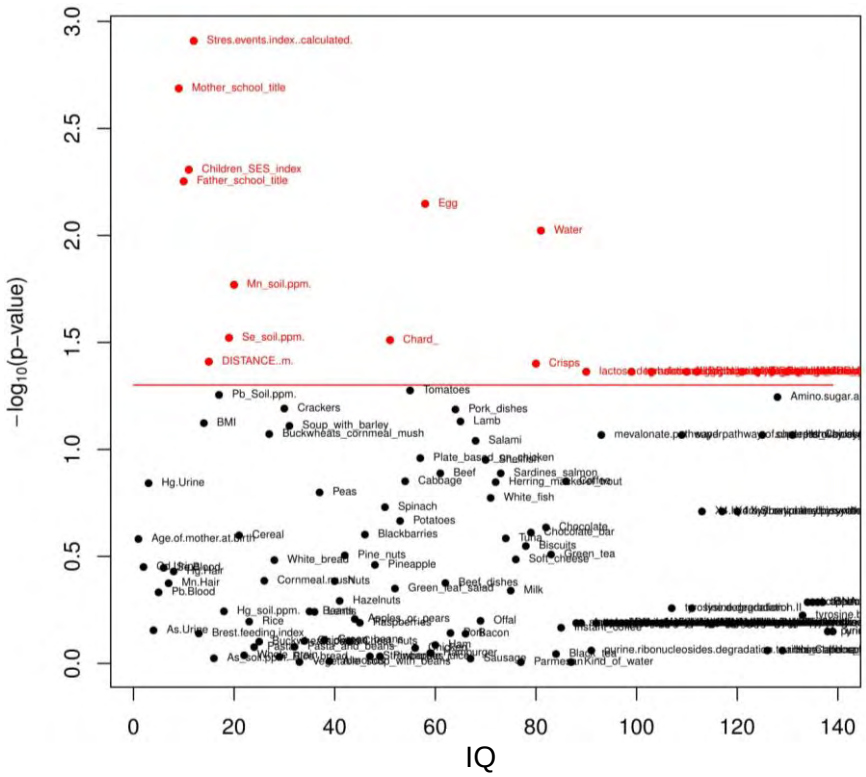


Figure 11.A: **Manhattan plot** showing the total IQ index with environmental, nutritional, demographic factors and pathways.

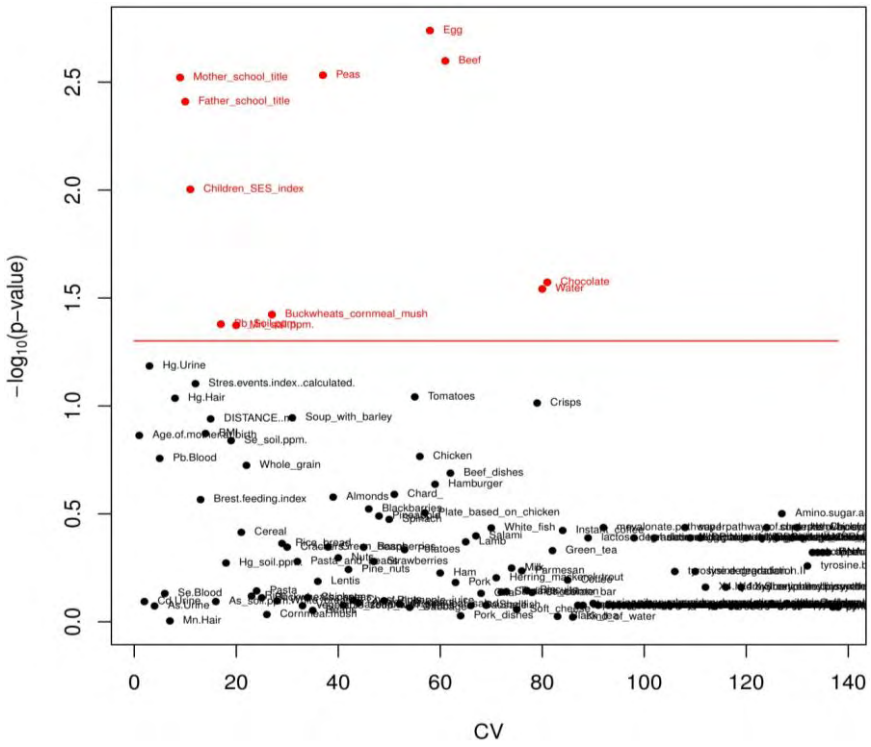


Figure 11.B: **Manhattan plot** showing the Verbal Comprehension Index with environmental, nutritional, demographic factors and pathways.

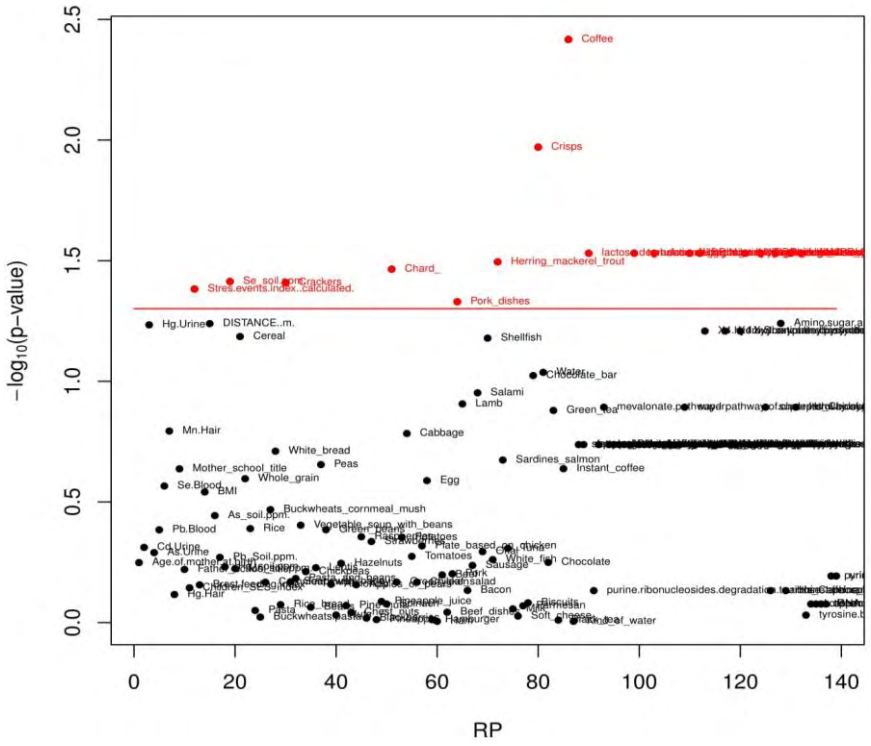


Figure 11.C: Manhattan plot showing the Visual Spatial index with environmental, nutritional, demographic factors and pathways.

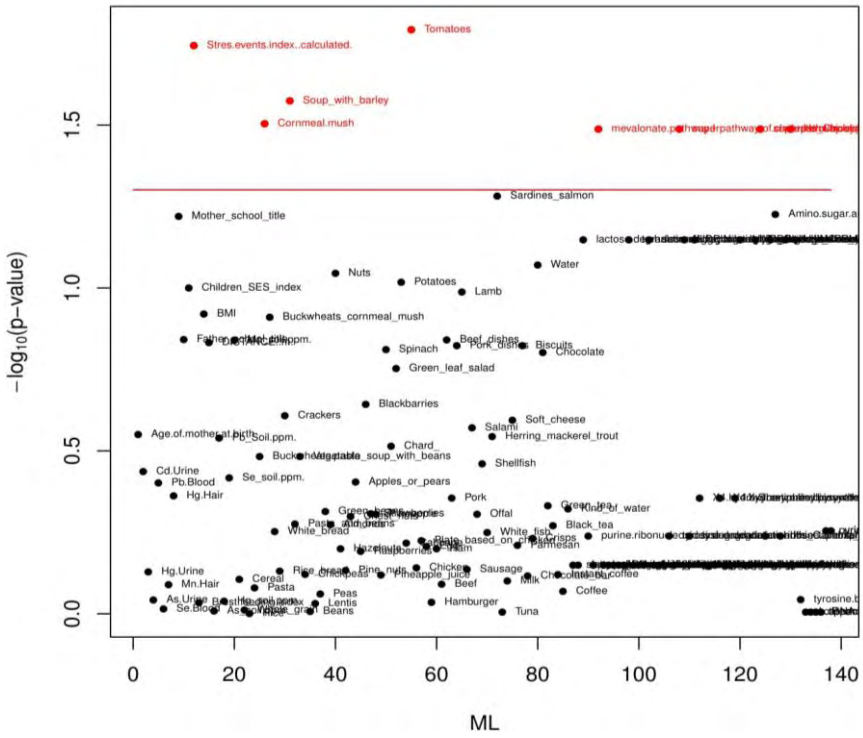


Figure 11.D: Manhattan plot showing the Working Memory index with environmental, nutritional, demographic factors and pathways.

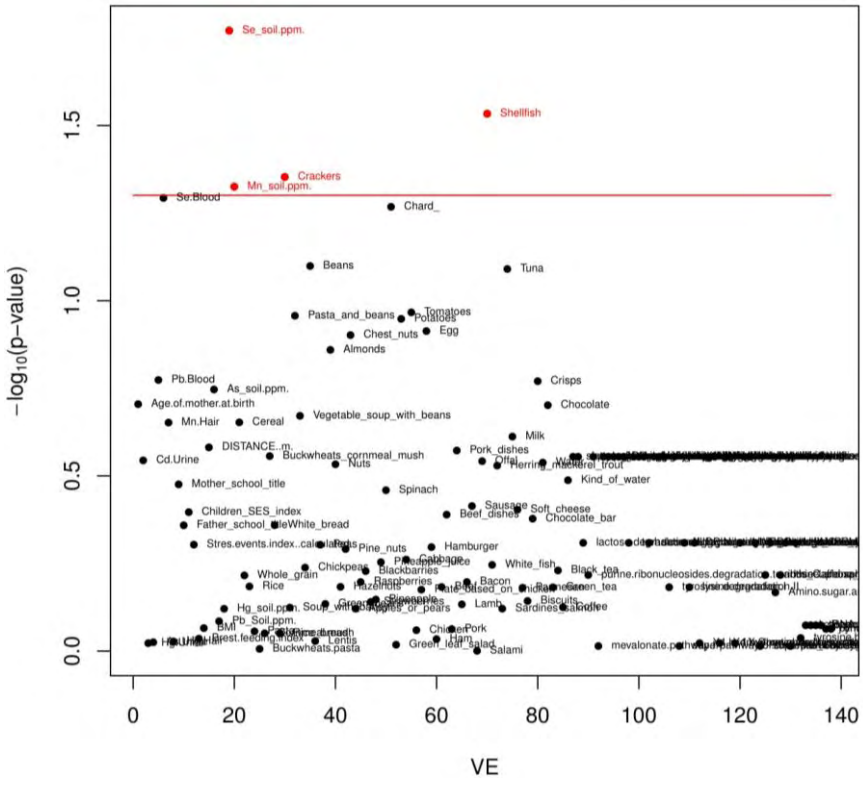


Figure 11.E: Manhattan plot showing the Processing Speed Index with environmental, nutritional, demographic factors and pathways.

EWAS analysis results relevant to the Total IQ index (IQ) are reported IN Figure 11.Afor the Manhattan plots and in Figure 12.Afor the Volcano maps. Manhattan plots show which factors are statistically significant for the respective health outcome. Volcano plots indicate not only statistical significance but also the sign of the association between exposure factors and metabolic pathways and the respective health outcome. The results for the Verbal Comprehension Index (CV), Visual Spatial Index (RP), Working Memory Index (ML) and Processing Speed Index (VE) are described using Manhattan plots (Figure 11.B, 11.C, 11.D, 11.E respectively) and using Volcano maps (Figure 12.B, 12B.C, 12.D, 12.E respectively).

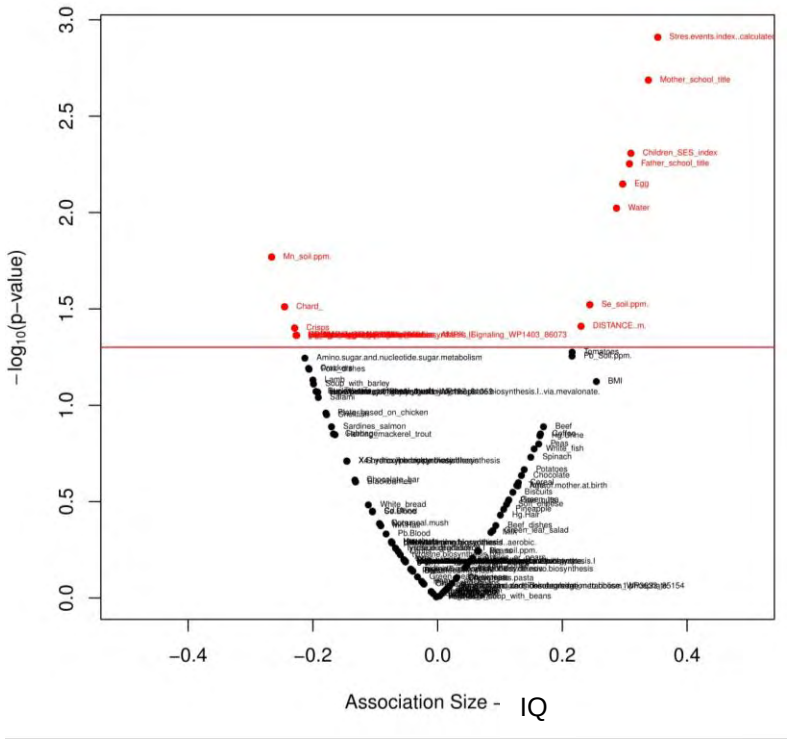


Figure 12.A: Volcano plot showing the total IQ index with environmental, nutritional, demographic factors and pathways.

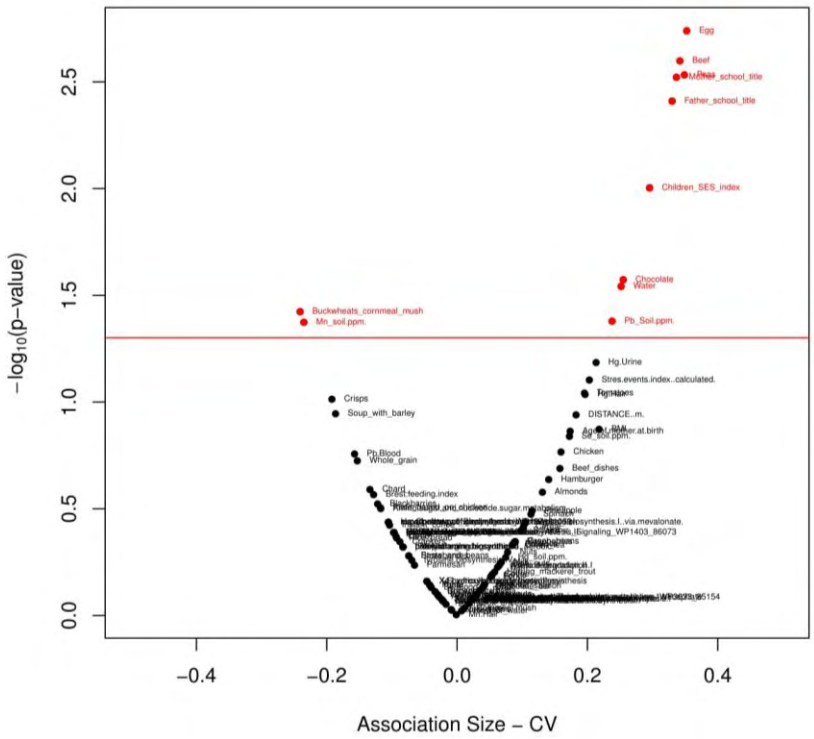


Figure 12.B: Volcano plot showing the Verbal Comprehension Index with environmental, nutritional, demographic factors and pathways.

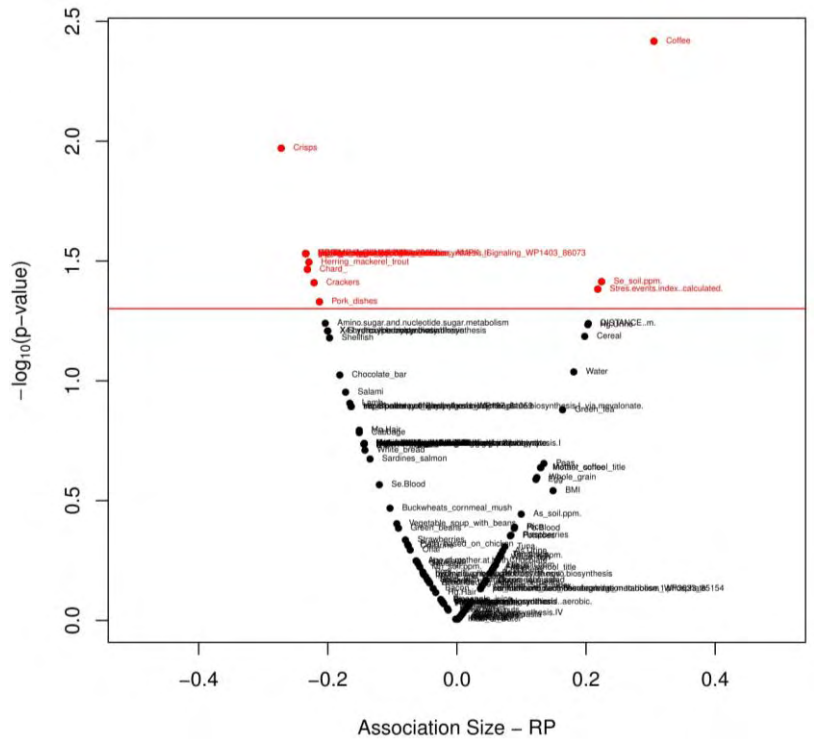


Figure 12.C: Volcano plot showing the Visual Spatial index with environmental, nutritional, demographic factors and pathways.

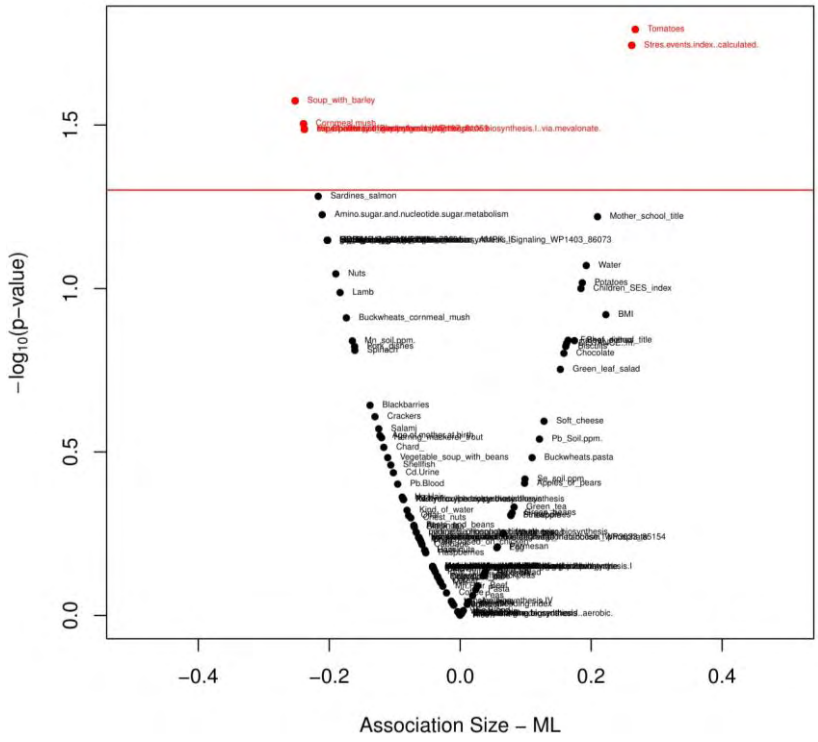


Figure 12.D: Volcano plot showing the Working Memory index with environmental, nutritional, demographic factors and pathways.

Figure 12.E: Volcano plot showing the Processing Speed Index with environmental, nutritional, demographic factors and pathways.

Results show that socio-cultural factors are strongly associated with IQ. More specifically *mother school title* and *father school title* show a robust statistical association (p-value<0.05) with most of the IQ indices considered. Looking at the volcano plots both parameters show a positive association with the IQ scores indicating that the higher educational level of the mothers and fathers may have positive impact on neurodevelopment of children. Of note, the parents' educational level influence the children's CV (verbal comprehension index) in the same positive manner.

The *stress events index* was derived by merging the total number of stressful events detected by the mother and their average intensity also plays an important role on the children's IQ, ML, RP (p-value<0.05) showing a positive association with the study factors. Stress events indices such as anxiety, decrease the levels of occurrence of neurodevelopmental disorders inasmuch, apart from IQ, enhances memory and visual spatial.

D*istance of the residence address of the family from the industrial plants* is a key factor associated with positive correlation with IQ. Analysis of the results show that living far from the industrial plants has a positive impact on the cognitive functions of the children.

Among the biomonitoring data *selenium* in blood shows a good statistical association (p-value <0.05) with VC, RP and IQ. Selenium is essential trace element uptaken up by humans through the food chain. This result confirms the antioxidant properties of selenium which is considered an important regulator of brain function providing protection from ROS-induced cell damage; the latter is strongly implicated in a number of neurologic disorders.

Interesting conclusions can be drawn from the analysis of food consumption patterns. *Tomato* consumption appears to be statistically (p-value <0.05) associated to ML index while *coffee* and *chocolate* consumption reveals a strong association with RP index and CV index respectively. Both these food items show a positive association meaning that their consumption has

potential positive effects on the cognitive functions of the children. Epidemiological evidence suggests that consumption of lycopene, natural antioxidant presents in tomatoes, is able to reduce the risk of chronic diseases such as cancer, cardiovascular diseases as well as psychiatric syndromes (Story et al., 2010). In another study (Li and Zhang, 2007) reported that low serum levels of lycopene have been associated with increased risk of psychiatric disorders.

Higher *consumption of bottled water and eggs* appears to have positive effects on the IQ and CV index, while higher consumption of crackers and other crisps and chard has a negative effect on the VE and RP, IQ and RP indices respectively. Shellfish has also a negative association with the VE index, presumably linked to the increased levels of neurotoxic metals.

Last but not least, *manganese in soil* appears to be inversely associated (p-value <0.05) with IQ, CV and VE. The negative sign of the association supports the positive impact of manganese on neurodevelopmental disorders.

In metabolic terms, the following pathways appear to associate negatively with IQ and RP. These pathways are: lactose degradation III, trehalose degradation II trehalase, sucrose degradation V mammalian, UDP N acetyl D galactosamine biosynthesis II, glycogen degradation III, glycogen degradation II, GDP glucose biosynthesis, Glycolysis Gluconeogenesis, Hs_AMPactivated_Protein_Kinase_.AMPK._Signaling_WP1403_86073, and Hs_Cori_Cycle_WP1946_79691. Four other metabolic pathways have negative effect on ML index: mevalonate pathway I, superpathway of cholesterol biosynthesis, superpathway of geranylgeranyldiphosphate biosynthesis I via mevalonate, Hs_Cholesterol_Biosynthesis_WP197_81059.

Discussion

The EWAS analysis performed on the Taranto cohort dataset revealed that sociodemographicfactors play a key role in the development of neurodevelopmental disorders in children as they are strongly statistically associated with a multitude of indices related to their cognitive function. These factors mostly include the maternal and paternal school title. Among the multitude of sociodemographic, nutrition and environmental factors evaluated, socioeconomic status seem to be dominant with regard to its effect on child neurodevelopmental disorders. This finding has also been established in other recent studies in the Italian population (Ronfani et al., 2015), strengthening the importance of socioeconomic status in affecting the different domains of early child development.

A further important parameter affecting the outcomes of many indices is the distance from the industrial plants. Also in this case the positive sign found in the association of residence address distance from industry with QI supports the assumption of a positive impact of living in the areas far from the industrial plants (Volcano plot, Figure 12.A). This result confirms that industry is a major pollution source for the population living around Taranto. In the area industrial activities are responsible for environmental contamination, mainly due to heavy metals, asbestos, polycyclic aromatic hydrocarbons (PAHs), organic solvents, polychlorinated biphenyls (PCBs) and dioxin (Marinaccio et al., 2011), which are transported via the air droplets -such as Windaham (2006) refers for autism spectrum disorders- or the food chain in children.

Human biomonitoring values measured in the children appear to be of comparable level with the results obtained from other studies in other sites and do not reveal any clear pattern associated with the residence of the children nor with the distance from the industrial plant. Overall the EWAS approach showed its powerful capability to unravel the associations between exposure determinants and neurodevelopmental disorders. The analysis performed considering all the potential dependent variables; however, it was hampered by the lack of information on environmental contamination which

would result in a more complete and insightful analysis that could be used more effectively for health risk management.

Among the environmental parameters, concentration of manganese appears to influence negatively many tests outcomes; higher soil concentration of manganeseis associated with worse performances in IQ, CV and VE test. Manganeseis absorbed by plant roots and transported directly to vegetarian man or indirectly through herbivores. These results are in full agreement with literature findings, which clearly indicate manganese as a neurotoxicant.

On the other hand, the positive effect of selenium is explained by the fact that selenium is an element providing protection from ROS-induced cell damage due to its antioxidant properties.

Analysis of the influence of the diet reveals some promising results such as the protective role of tomatoes or the negative effect of crackers, chards and crisps. Overall, results are sometimes discordant, requiring further analysis, since firm conclusions are hampered by the lack of information about the content of metals in the food items considering as well the cross-sectional character of the study.

Shellfish consumption was identified as another key determinant for VE. It seems that consumption of shellfish led to human uptake of a significant dose of heavy metals, which bioaccumulated in shellfish provoking neurological disorders(Awada and Kojan, 2003).

Among the various pathways identified, the following ones have been associated to lower scores of neurological development: lactose degradation III, trehalose degradation II trehalase, sucrose degradation V for mammalians,UDP N acetyl D galactosamine biosynthesis II, glycogen degradation III, glycogen degradation II, GDP glucose biosynthesis,GlycolysisGluconeogenesis,Hs_AMPactivated_Protein_Kinase_. AMPK._ Signaling_WP1403_86073, and Hs_Cori_Cycle_WP1946_79691,

Mevalonate pathway I, superpathway of cholesterol, superpathway of geranylgeranyldiphosphate biosynthesis I via mevalonate.

The mevalonate cascade is a key metabolic pathway that regulates a variety of cellular functions and is thereby involved in the pathophysiology of many brain diseases, including neurodevelopmental and neurodegenerative disorders (Jiao et al., 2016). Mevalonate pathway correlated with both cholesterol and geranylgeranyldiphosphate biosynthesis pathways, resulting incentral nervous system deregulation. For example, the key biochemicalfinding for mevalonate kinase deficiency is the accumulation of urinary mevalonic acid / mevalonolactone(Kanungo et al., 2013). Moreover the 7-dehydrolesterol, the last compound before the cholesterol biosynthesis, is involved in the vitamin D_3 biosynthesis (Figure 13). The no activation of the pathway blocks the D_3 biosynthesis, which has associated with neurodevelopmental disorders. According to Kocovska et al. (2012), Vitamin D deficiency may be an environmental trigger for ASD in individuals genetically predisposed for the broad phenotype of autism and the recognition of this possibly important role of vitamin D in ASD (Kocovska et al., 2012).

Several indirect evidences exist that autophagy may degrade the glycogen in neurons, in a way similar to many other tissues. For example, defects in lysosomal enzyme acid alpha-glucosidase have been associated to glycogen accumulation in many tissues, including neurons (Sidman et al., 2008);the concentration of glycogen in neurons provokes neurodegenaration.

Trehalose acts as a potent stabilizer of proteins able to preserve protein structural integrity and reduces aggregation of pathologically misfolded proteins (Emanuele, 2014). Trehalose is an mTOR-independent inducer of autophagy and together with TSC, form a complex which is the key regulator of protein synthesis, therefore, controlling cell growth (Zhou and Parada, 2012). Recent data suggest that the TSC1/2 complex influences also neural polarity.

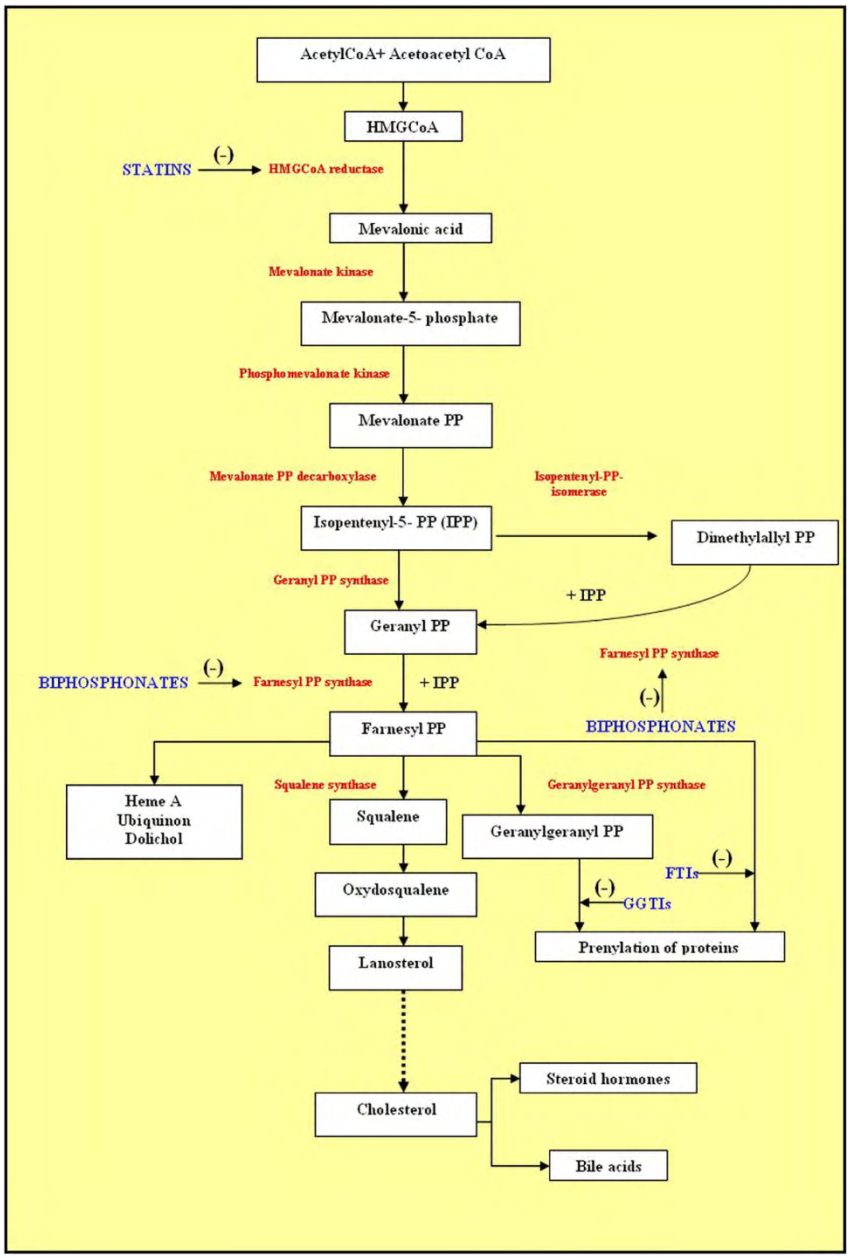


Figure 13: The pathway of mevalonate acid.

Analysis and results of these associations could be greatly improved by the availability of environmental pollution data (soil, air and water) which can allow gathering further insights into the potential associations between environmental contamination, diet and exposure. Moreover, chemical analysis of food items consumed by the local population would greatly improve the association between exposure through diet and neurodevelopmental disorders observed.

Conclusions

In this study, an advanced statistical modeling approach (EWAS) has been employed aiming at the analysis of potential associations between exposure and exposure modifiers on the neurodevelopmental cognitive functions of children living around Taranto (Italy) which is characterized by the presence of the largest steel transformation and production plant of Europe with a capacity to produce 10 million tons of steel annually, corresponding to 40% of the Italian steel production.

Our analysis identified that the most influential parameters are the socio-economic-cultural conditions of the family. The positive effect of these factors on development of cognitive functions in children is associated to the better quality of life, translated in terms of nutrition of higher quality, lower exposure to environmental contamination and better educational activities and opportunities. Similarly, parental school title has been associated positively with cognitive development.

Beyond the effect of these factors, exposure related parameters seem to be important for specific neurodevelopmental tests. Distance from the industrial plants affects the outcomes of many indices of the four neurodevelopmental test batteries analyzed impacting negatively the population living in the areas closer to the industrial plants, confirms the hypothesis that steel industry is a major pollution source for the population living in the area of Taranto, eventually with additional pollutants than the ones described in the biomonitoring data.

Although exposure to heavy metals for the children of Taranto appear to be of comparable level with studies performed in other sites, exposure to heavy metals seems to affect specific cognitive functions. Higher concentration of manganese is negatively associated with neurodevelopmental progress, while it has a clear effect on IQ, CV and VE; presence of selenium appears to produce positive effects, due to its antioxidant properties.

Among the investigated parameters, specific dietary components seem to have either a positive or a negative influence, however, due to the lack of food residue data the associations are many times inconclusive. However, it is clearly demonstrated that foods with antioxidants (e.g. tomatoes) have a clearly beneficial effect on cognitive functions.

Finally, induction of key metabolic pathways (e.g. lactose degradation III and sucrose degradation V for mammalian cells) have been negatively associated with neurodevelopmental progress.

Considering the interesting findings highlighted above, it is recommended that the analysis should be repeated considering additional environmental contamination data from multiple media (soil, air, water and eventually food residues). Moreover, considering the impact of organic chemical compounds on neurodevelopmental disorders such as PCBs and dioxins the study will greatly benefit from availability of both HBM and environmental data relevant to these compounds. Moreover, incorporation of pollutant toxicokinetics analysis (in addition to exposure data) will provide additional insights regarding the actual internal and biologically effective dose of these xenobiotics in the brain.

Abbreviations

LC	Liquid chromatography
MS	Mass spectrometry
GC	Gas chromatography
ASD	Autism Spectrum Disorder
HMDB	Human Metabolome Database
KEGG	Kyoto Encyclopedia of Genes and Genomes
UPLC-MS	Ultra performance liquid chromatography – tandem mass spectrometry
ESI	Electrospray ionization
SES	Socioeconomic Status
QC	Quality control
MPP	Mass Profiler Professional
EWAS	Environment-wide association study
IQ	Total IQ index
CV	VerbalComprehensionIndex
RP	VisualSpatialIndex
ML	WorkingMemoryIndex
VE	ProcessingSpeedIndex

Bibliography

Awada A, Kojan S. Neurological disorders and travel. International journal of antimicrobial agents 2003; 21: 189-192.

Bisceglia L, de Nichilo G, Elia G, Schiavulli N, Minerba A, Greco L, Assennato G. [Assessment of occupational exposure to PAH in coke-oven workers of Taranto steel plant through biological monitoring]. Epidemiol Prev 2005; 29: 37-41.

Broix L, Jagline H, L Ivanova E, Schmucker S, Drouot N, Clayton-Smith J, Pagnamenta AT, Metcalfe KA, Isidor B, Louvier UW, Poduri A, Taylor JC, Tilly P, Poirier K, Saillour Y, Lebrun N, Stemmelen T, Rudolf G, Muraca G, Saintpierre B, Elmorjani A, Deciphering Developmental Disorders s, Moise M, Weirauch NB, Guerrini R, Boland A, Olaso R, Masson C, Tripathy R, Keays D, Beldjord C, Nguyen L, Godin J, Kini U, Nischke P, Deleuze J-F, Bahi-Buisson N, Sumara I, Hinckelmann M-V, Chelly J. Mutations in the HECT domain of NEDD4L lead to AKT-mTOR pathway deregulation and cause periventricular nodular heterotopia. Nat Genet 2016; advance online publication.

Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput 2000: 418-29.

Bystritsky A, Khalsa SS, Cameron ME, Schiffman J. Current diagnosis and treatment of anxiety disorders. P T 2013; 38: 30-57.

Caliendo AM, Gilbert DN, Ginocchio CC, Hanson KE, May L, Quinn TC, Tenover FC, Alland D, Blaschke AJ, Bonomo RA, Carroll KC, Ferraro MJ, Hirschhorn LR, Joseph WP, Karchmer T, MacIntyre AT, Reller LB, Jackson AF, Infectious Diseases Society of A. Better tests, better care: improved diagnostics for infectious diseases. Clin Infect Dis 2013; 57 Suppl 3: S139-70.

Caubit X, Gubellini P, Andrieux J, Roubertoux PL, Metwaly M, Jacq B, Fatmi A, Had-Aissouni L, Kwan KY, Salin P, Carlier M, Lieden A, Rudd E, Shinawi M, Vincent-Delorme C, Cuisset J-M, Lemaitre M-P, Abderrehamane F, Duban B, Lemaitre J-F, Woolf AS, Bockenhauer D, Severac D, Dubois E, Zhu Y, Sestan N, Garratt AN, Le Goff LK, Fasano L. TSHZ3 deletion causes an

autism syndrome and defects in cortical projection neurons. Nat Genet 2016; advance online publication.

Chen A, Dietrich KN, Huo X, Ho SM. Developmental neurotoxicants in e-waste: An emerging health concern. Environmental Health Perspectives 2011; 119: 431-438.

Dervola KSN, Johansen EB, Walaas SI, Fonnum F. Gender-dependent and genotype-sensitive monoaminergic changes induced by polychlorinated biphenyl 153 in the rat brain. NeuroToxicology 2015; 50: 38-45.

Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev 2007; 26: 51-78.

Durak O, Gao F, Kaeser-Woo YJ, Rueda R, Martorell AJ, Nott A, Liu CY, Watson LA, Tsai L-H. Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. Nat Neurosci 2016; advance online publication.

Emanuele E. Can trehalose prevent neurodegeneration? Insights from experimental studies. Curr Drug Targets 2014; 15: 551-7.

Emond P, Mavel S, Aïdoud N, Nadal-Desbarats L, Montigny F, Bonnet-Brilhault F, Barthélémy C, Merten M, Sarda P, Laumonnier F, Vourc'h P, Blasco H, Andres CR. GC-MS-based urine metabolic profiling of autism spectrum disorders. Analytical and Bioanalytical Chemistry 2013; 405: 5291-5300.

Eyles JB, J. "Environments, Risks and Health: Social Perspectives": Routledge, 2016.

Feigin A. Evidence from biomarkers and surrogate endpoints. NeuroRx 2004; 1: 323-30.

Frumkin H. "Environmental Health: From Global to Local". Vol 2: Wiley, 2010.

Gerhold DL, Jensen RV, Gullans SR. Better therapeutics through microarrays. Nat Genet 2002; 32 Suppl: 547-51.

Giulivi C, Napoli E, Tassone F, Halmai J, Hagerman R. Plasma Biomarkers for Monitoring Brain Pathophysiology in FMR1 Premutation Carriers. Front Mol Neurosci 2016; 9: 71.

Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, West AB, Arking DE. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nature Communications 2014; 5: 5748.

Hall MA, Dudek SM, Goodloe R, Crawford DC, Pendergrass SA, Peissig P, Brilliant M, McCarty CA, Ritchie MD. Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield Personalized Medicine Research Project Biobank. Pac Symp Biocomput 2014: 200-11.

Hawi Z, Cummins TD, Tong J. Rare DNA variants in the brain-derived neurotrophic factor gene increase risk for attention-deficit hyperactivity disorder: a next-generation sequencing study. 2016.

Heberling C, Dhurjati P. Novel systems modeling methodology in comparative microbial metabolomics: identifying key enzymes and metabolites implicated in autism spectrum disorders. Int J Mol Sci 2015; 16: 8949-67.

Horgan RP, Kenny LC. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. J Obstet Gynaecol. 2011: 189-95.

Inoue YU, Inoue T. Brain enhancer activities at the gene-poor 5p14.1 autism-associated locus. Scientific Reports 2016; 6: 31227.

Ioannidis JP, Loy EY, Poulton R, Chia KS. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. Sci Transl Med 2009; 1: 7ps8.

Jiao X, Ashtari N, Balaei MR, Chen QM, Badbezanchi I, Shojaei S, Marzban A, Mirzaei N, Chung S, Guan T, Li J, Vriend J, Mehr SE, Kong J, Marzban H. Mevalonate Cascade and Neurodevelopmental and Neurodegenerative Diseases: Future Targets for Therapeutic Application. Curr Mol Pharmacol 2016.

Kanungo S, Soares N, He M, Steiner RD. Sterol metabolism disorders and neurodevelopment-an update. Dev Disabil Res Rev 2013; 17: 197-210.

Katajamaa M, Miettinen J, Oresic M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 2006a; 22: 634-6.

Katajamaa M, Miettinen J, Orešič M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 2006b; 22: 634-636.

Kell DB. Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug Discov Today 2006; 11: 1085-92.

Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays 2004; 26: 99-105.

Kobayashi T, Nishiumi S, Ikeda A, Yoshie T, Sakai A, Matsubara A, Izumi Y, Tsumura H, Tsuda M, Nishisaki H, Hayashi N, Kawano S, Fujiwara Y, Minami H, Takenawa T, Azuma T, Yoshida M. A novel serum metabolomics-based diagnostic approach to pancreatic cancer. Cancer Epidemiol Biomarkers Prev 2013; 22: 571-9.

Kocovska E, Fernell E, Billstedt E, Minnis H, Gillberg C. Vitamin D and autism: clinical review. Res Dev Disabil 2012; 33: 1541-50.

Lei Z, Huhman DV, Sumner LW. Mass spectrometry strategies in metabolomics. J Biol Chem 2011; 286: 25435-42.

Lelieveld SH, Reijnders MRF, Pfundt R, Yntema HG, Kamsteeg E-J, de Vries P, de Vries BBA, Willemsen MH, Kleefstra T, Lohner K, Vreeburg M, Stevens SJC, van der Burgt I, Bongers EMHF, Stegmann APA, Rump P, Rinne T, Nelen MR, Veltman JA, Vissers LELM, Brunner HG, Gilissen C. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat Neurosci 2016; 19: 1194-1196.

Li Y, Zhang J. Serum concentrations of antioxidant vitamins and carotenoids are low in individuals with a history of attempted suicide. Nutr Neurosci 2007; 10: 51-8.

Mandel SA, Morelli M, Halperin I, Korczyn AD. Biomarkers for prediction and targeted prevention of Alzheimer's and Parkinson's diseases: evaluation of drug clinical efficacy. EPMA J 2010; 1: 273-92.

Marinaccio A, Belli S, Binazzi A, Scarselli A, Massari S, Bruni A, Conversano M, Crosignani P, Minerba A, Zona A, Comba P. Residential proximity to industrial sites in the area of Taranto (Southern Italy). A case-control cancer incidence study. Ann Ist Super Sanita 2011; 47: 192-9.

Mavel S, Nadal-Desbarats L, Blasco H, Bonnet-Brilhault F, Barthélémy C, Montigny F, Sarda P, Laumonnier F, Vourc′ h P, Andres CR, Emond P. 1H–13C NMR-based urine metabolic profiling in autism spectrum disorders. Talanta 2013; 114: 95-102.

Nordstrom A, Want E, Northen T, Lehtio J, Siuzdak G. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. Anal Chem 2008; 80: 421-9.

Noto A, Fanos V, Barberini L, Grapov D, Fattuoni C, Zaffanello M, Casanova A, Fenu G, De Giacomo A, De Angelis M, Moretti C, Papoff P, Ditonno R, Francavilla R. The urinary metabolomics profile of an Italian autistic children population and their unaffected siblings. J Matern Fetal Neonatal Med 2014; 27 Suppl 2: 46-52.

Novice R. "Overview of the environment and health in Europe in the 1990s.". WHO, 1999.

Palmer JA, Poenitzsch AM, Smith SM, Conard KR, West PR, Cezar GG. Metabolic Biomarkers of Prenatal Alcohol Exposure in Human Embryonic Stem Cell-derived Neural Lineages. Alcoholism, clinical and experimental research 2012; 36: 1314-1324.

Patel CJ, Ioannidis JP. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. J Epidemiol Community Health 2014; 68: 1096-100.

Patel CJ, Manrai AK. Development of exposome correlation globes to map out environment-wide associations. Pac Symp Biocomput 2015: 231-42.

Phadikar P, Saxena S, Ruia S, Lai TY, Meyer CH, Eliott D. The potential of spectral domain optical coherence tomography imaging based retinal biomarkers. Int J Retina Vitreous 2017; 3: 1.

Putschögl FM, Gaum PM, Schettgen T, Kraus T, Gube M, Lang J. Effects of occupational exposure to polychlorinated biphenyls on urinary metabolites of neurotransmitters: A cross-sectional and longitudinal perspective. International Journal of Hygiene and Environmental Health 2015; 218: 452-460.

Ronfani L, Vecchi Brumatti L, Mariuz M, Tognin V, Bin M, Ferluga V, Knowles A, Montico M, Barbone F. The Complex Interaction between Home Environment, Socioeconomic Status, Maternal IQ and Early Child Neurocognitive Development: A Multivariate Analysis of Data Collected in a Newborn Cohort Study. PLoS One 2015; 10: e0127052.

Sidman RL, Taksir T, Fidler J, Zhao M, Dodge JC, Passini MA, Raben N, Thurberg BL, Cheng SH, Shihabuddin LS. Temporal neuropathologic and behavioral phenotype of 6neo/6neo Pompe disease mice. J Neuropathol Exp Neurol 2008; 67: 803-18.

Sparkman OD. Review of the 48th ASMS conference on mass spectrometry and allied topics held in Long Beach, California June 11-15, 2000. J Am Soc Mass Spectrom 2000; 11: 921.

Story EN, Kopec RE, Schwartz SJ, Harris GK. An update on the health effects of tomato lycopene. Annu Rev Food Sci Technol 2010; 1: 189-210.

Strachan T. Genetics and Genomics in Medicine, 2017.

Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS 2010; 5: 463-6.

Tan KC, Ipcho SV, Trengove RD, Oliver RP, Solomon PS. Assessing the impact of transcriptomics, proteomics and metabolomics on fungal phytopathology. Mol Plant Pathol 2009; 10: 703-15.

Tuck MK, Chan DW, Chia D, Godwin AK, Grizzle WE, Krueger KE, Rom W, Sanda M, Sorbara L, Stass S, Wang W, Brenner DE. Standard operating procedures for serum and plasma collection: early detection research network

consensus statement standard operating procedure integration working group. J Proteome Res 2009; 8: 113-7.

Viviano G, Ziemacki G, Settimo G, Cattani G, Spartera M, Catucci F, Carbotti G. [Air quality assessment in an urban-industrial area: the Taranto case study]. Epidemiol Prev 2005; 29: 45-9.

Wang H, Liang S, Wang M, Gao J, Sun C, Wang J, Xia W, Wu S, Sumner SJ, Zhang F, Sun C, Wu L. Potential serum biomarkers from a metabolomics study of autism. J Psychiatry Neurosci 2016; 41: 27-37.

Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, Holmes E, Nicholson JK. Global metabolic profiling procedures for urine using UPLC-MS. Nat Protoc 2010; 5: 1005-18.

Wechsler D. Wechsler Intelligence Scale for Children. Psychological Corporation, San Antonio TX, 2003.

West PR, Amaral DG, Bais P, Smith AM, Egnash LA, Ross ME, Palmer JA, Fontaine BR, Conard KR, Corbett BA, Cezar GG, Donley EL, Burrier RE. Metabolomics as a tool for discovery of biomarkers of autism spectrum disorder in the blood plasma of children. PLoS One 2014; 9: e112445.

Westerhoff HV, Palsson BO. The evolution of molecular biology into systems biology. Nat Biotechnol 2004; 22: 1249-52.

WHO. "Health topics: Environmental health", 2015.

Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 2005; 14: 1847-50.

Windham GC, Zhang L, Gunier R, Croen LA, Grether JK. Autism spectrum disorders in relation to distribution of hazardous air pollutants in the San Francisco Bay area. Environmental health perspectives 2006: 1438-1444.

Wu YE, Parikshak NN, Belgard TG, Geschwind DH. Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. Nat Neurosci 2016; advance online publication.

Yin P, Lehmann R, Xu G. Effects of pre-analytical processes on blood samples used in metabolomics studies. Anal Bioanal Chem 2015; 407: 4879-92.

Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC bioinformatics 2013; 14: 1.

Zhou B, Wang J, Ressom HW. MetaboSearch: tool for mass-based metabolite identification using multiple databases. PLoS One 2012; 7: e40096.

Zhou J, Parada LF. PTEN signaling in autism spectrum disorders. Curr Opin Neurobiol 2012; 22: 873-9.